Standing Panel on Impact Assessment

Remote Sensing for Impact Evaluation of Agriculture and Natural Resource Management Research: Guidelines for Use in One CGIAR

Johanne Pelletier, Casey Maue, Mina Karasalo, Kelsey Jack, Julio Barros

August 2023

Citation: Pelletier, J., Maue, C., Karasalo, M., Jack, K., Barros, J. (2023). ***Remote sensing for impact evaluation of agriculture and natural resource management research: Guidelines for use in One CGIAR*****. Rome: Standing Panel on Impact Assessment (SPIA).**

Design and layout: Luca Pierotti and Macaroni Bros

Remote Sensing for Impact Evaluation of Agriculture and Natural Resource Management Research: Guidelines for Use in One CGIAR

Johanne Pelletier, Casey Maue, Mina Karasalo, Kelsey Jack, Julio Barros

August 2023

Contents

Executive Summary

Remote sensing is developing at a rapid pace, with satellite-based Earth observation (EO) data being made available freely, openly, and at higher spatial, temporal, and spectral resolutions than ever before. This provides new opportunities for EO data to complement traditional survey data and improve the rigor and scope of impact evaluations. This document provides methodological guidance for the use of EO data in measuring the impacts of innovations and interventions on outcomes related to agriculture, natural resources, livestock, and the environment. Though the particular aim is to support incorporation of EO data into evaluation of One CGIAR activities, these guidelines are also relevant for a broad audience of researchers and practitioners from a variety of backgrounds interested in learning about the potential for using remote sensing to conduct impact evaluation.

We showcase 11 case studies that use EO data for impact evaluation. These case studies—representing primarily a set of recent and ongoing projects sponsored by the CGIAR—highlight the range of use cases for EO data as well as several innovative methodological approaches to their analysis. Our discussion of these case studies draws from interviews with members of the implementing research teams. The lessons these researchers learned from their experiences demonstrate how the challenges of using EO data for impact evaluation can be identified and addressed in a variety of real-world research contexts.

We provide strategic insights arising from this portfolio of case studies to help identify in which context remote sensing data can be relevant and appropriate for impact evaluation. We reflect on the skill set, background, and experience that can make up a successful research team that use EO for impact evaluation.

The core of these guidelines centers on introducing a set of good practices for the appropriate use of EO data in impact evaluation. We highlight key topics that must be considered by research teams seeking to use EO data, including image availability, image pre-processing, reference data collection, and validation. In addition, we discuss how issues related to scale, spatial correlation, and nonclassical measurement error can arise when combining EO and other types of spatial data. We provide insights on the transfer of existing geospatial approaches to other contexts.

Echoing the new global impetus for monitoring and measuring progress toward the attainment of the Sustainable Development Goals (SDGs) using EO, we assess and describe how the sustainable development targets identified by One CGIAR initiative proposals can be related to EO-based indicators. For each target, we identify relevant EO-based variables, the recent literature demonstrating their utility, and resources for accessing satellite-derived data sets or products.

We close with a discussion of the need for a systematic change in data collection to fully leverage the potential of EO data for measuring impacts of activities within One CGIAR.

1. Introduction

In this time of planetary crisis, it is more urgent than ever that policy interventions produce expected outcomes. The global sustainability challenge is immense, demanding a systemic transformation of food, land, and water systems in order to meet United Nations Sustainable Development Goal (SDG) targets (CGIAR System Organization, 2021). Addressing the climate crisis head on, the CGIAR's new 2030 research and innovation strategy strives to deliver science and innovation that can advance this transformation. Measuring the impacts of this transformative agenda is imperative. As policymakers around the world strive toward increasingly challenging and complex global objectives—often with limited resources—the demand for evidence-based policymaking heightens the importance of tools that can facilitate rigorous impact evaluations at scale under often-short policy timelines.

One CGIAR's mission:
To deliver science and innovation that advance the transformation of food, land, and water systems in a climate crisis.

Causal impact evaluation (IE) is a core part of the broader agenda of evidence-based policymaking and can play a pivotal role in learning from interventions and guiding actions based on past experiences. Recent calls for more rigorous evaluation of impacts from programs in several branches of research and development (Stevenson et al., 2018) stem from the need to invest finite resources effectively.

Satellite-based Earth observation (EO) data are measures of the physical environment at the Earth's surface recorded by satellites orbiting the Earth. EO data are being made available freely, openly, and at higher spatial, temporal, and spectral resolutions than ever before. The increasing availability of EO data, in conjunction with improvements in data management, processing, and analysis technologies, brings new and exciting possibilities for measuring the impacts of research and innovations within One CGIAR. Especially in developing countries, where other data sources are often unavailable, EO data have the potential to support impact evaluations in settings where they would otherwise be practically infeasible. Despite their promise, however, EO data do not solve all challenges related to impact evaluation. Indeed, they introduce new issues that must be addressed for impact evaluations based upon these data to be meaningful.

To date, the use of remotely sensed data to evaluate the impact of CGIAR innovations has been limited (Stevenson et al., 2018). Promoting their effective use more broadly requires that researchers and practitioners understand new concepts and requirements underlying the use of EO for impact evaluation. This work provides practical guidance for future research projects that seek to use remotely sensed data to measure the impacts of CGIAR innovations on outcomes related to agriculture, natural resources, livestock, and the environment. We identify appropriate opportunities and common pitfalls, as well as the potential advantages and challenges associated with the rigorous use of EO data to measure adoption and/or outcomes of technologies. We draw from recent and ongoing projects sponsored by CGIAR that have incorporated EO data into impact evaluations to highlight a variety of real-world applications and summarize the lessons learned by its research teams.

This document proposes a nontechnical overview of geospatial impact evaluation methods. It is appropriate for researchers and practitioners from a diverse array of backgrounds, including those in the remote sensing community, as well as social and biophysical scientists. We assume no prior expertise in remote sensing, though technical concepts from remote sensing science, economics, and statistics will be discussed. This document is not intended to replace formal training in remote sensing science, the economics of impact evaluation, or the statistics of causal inference. Readers seeking to review concepts related to impact evaluation or remote sensing should refer to Annex 1 or Annex 2, respectively. We encourage readers to push their learning journey beyond these guidelines by supplementing ample references and links to online materials on data providers, existing products, satellite sensors, and online tutorials in Annex 5, 6, 7, and 8, respectively.

Ultimately, we hope these guidelines can make geospatial impact evaluation approaches more accessible and support the integration of remote sensing technologies into the impact evaluation methodologies deployed throughout One CGIAR. We hope they can promote better collaboration between the remote sensing science and impact evaluation communities within and beyond the CGIAR system.

In the following section, we describe the method used to develop these guidelines (section 2). Then, we describe the primary use for EO data in the context of impact evaluation and highlight key criteria for using them effectively (section 3). Next, we discuss the advantages of remote sensing technology for impact evaluation (section 4) and offer reflections on building effective geospatial impact evaluation research teams (section 5). Then, we provide methodological guidance on key issues for using remote sensing data for impact evaluation (section 6). We report on the state of the art EO applications relevant to One CGIAR impact areas (section 7). Section 8 concludes with a call for a change in data collection and on the strategic role that One CGIAR can play to leverage the full benefits of EO for impact evaluation. The reader will also find a description of each case study in Annex 3 and the main lessons learned from these case studies in Annex 4. Annex 5 provides a list of common data providers where users can access EO data from different satellites. Annex 6 provides information on EO products, including off-the-shelf products, that are directly relevant to One CGIAR activities. Annex 7 provides the characteristics of common satellite imagery programs that may be relevant to geospatial impact evaluation. Annex 8 offers a list of online training and tutorials for some of the most common open source remote sensing software.[1] Throughout the main text, we denoted with a (R) for 'review' and a blue background the sections that contain a review for readers with a remote sensing background, but consist of more advanced technical notes on remote sensing concepts for a broad audience.

2. Approach to producing these guidelines

These guidelines were developed from three primary sources of information. First, we conducted a review of foundational books and articles in the remote sensing literature to identify best practices and challenges associated with using satellite EO data. Second, we analyzed 11 case studies of projects recently funded by SPIA to gather input on researchers' applications of EO for impact evaluation. Third, we evaluated which outcome variables targeted by recent CGIAR initiatives could be measured by EO, given the state of the art in remote sensing applications, to ensure that these guidelines are relevant for future One CGIAR research objectives.

The case studies we reviewed were originally selected for funding because of the high quality of their proposed impact evaluation methods. At the time of this writing, some projects have been completed, while others are still in progress; we indicate in the text the status of completion of each project. They generally focus on evaluating the impact of a single CGIAR innovation in a single country, several years after the innovation was initially distributed and adopted. They vary in terms of the type of satellite imagery and remote sensing methods they relied on, the innovation they evaluated, and whether they use remote sensing to measure technology adoption, outcomes of interest, or both. The key aspects of each case study are described in Annex 3 (Table S1). For each case, we conducted semi-structured virtual interviews with members of each project, based on a set of questions targeted to their specific context and application of remote sensing. Interviews were recorded and transcribed for analysis. We also reviewed and analyzed all technical documentation related to each project, including proposals, reports, and related publications.

For our assessment of outcomes relevant to CGIAR objectives, we first reviewed recent CGIAR initiative proposals that explicitly mentioned "remote sensing," "spatial data", and/or "satellite imagery" to identify their specific sustainable development targets and how they relate to a set of key outcome variables. We then added to this initial set to cover all of CGIAR's five main impact areas. We analyzed the overlap between these targets and existing EO-based measurements, considering the level of readiness of the latter for impact evaluation.

[1] Readers may also consult the list of acronyms (section 11) and the glossary of terms (section 12).

3. How and when can remotely sensed Earth observation data be used for impact evaluation?

The goal of a causal impact evaluation in the CGIAR context is to measure adoption of a particular innovation and to quantify how this adoption affects processes or outcomes (e.g., poverty, yield, greenhouse gas emissions) within some study population over some period of time.[2] In this document, we use the term "technology" to refer to an innovation (or bundle of innovations) that affect processes or outcomes in agricultural or natural resource management settings. It refers to physical goods, such as improved seeds or fertilizer, as well as nonphysical inputs, such as management practices or policy interventions. To implement a causal impact evaluation, researchers must obtain data on both adoption of the innovation and on outcomes. Traditionally, these data are collected with field-based methods, through direct observations of a newly introduced technology adoption and outcomes, using household surveys, for instance.

EO data can be used to measure technology adoption (or more broadly the take up of an innovation), outcomes, or both, when there is a spatially-explicit identifier for the innovation that is being evaluated. Sometimes information about adoption is provided by an existing dataset of technology adoption that can be linked to existing spatial information. For instance, when a program is rolled out in a subset of districts or villages, adoption status can be linked to preexisting spatial files of district boundaries (polygons) or villages (points or polygons of boundaries). In this case, the spatial data are created by merging the adoption dataset and the spatial administrative boundary data sets (spatial units of observations), which can be used to look at the outcomes with remote sensing.

Satellite images can be analyzed to produce two main types of modeled outputs or maps: a categorical variable by classification or a continuous variable by regression.[3] Classification analysis involves simplifying landscape characteristics into a set of discrete classes and is commonly used in remote sensing to create, for example, thematic maps of land cover, land-cover change, or crop types. Regression analysis is used when the output maps a continuous variable, such as percentage of tree cover, biomass, crop yield, or soil carbon density. In general, when using EO, technology adoption is mapped using classification because it is a binary (adopters versus non-adopters) or categorical variable. In contrast, outcomes are typically mapped using regression, because we are interested in measuring the change in a continuous variable. For measuring both adoption and outcomes, it is essential to have accurate information about ground conditions. This information, also known as reference data, refers to georeferenced information (points or polygons) about ground conditions that is used for training and validation of EO-based classification and/or regression maps. We discuss this topic further in section 5d.

Measuring adoption with EO

Typically, to identify who has adopted an innovation, one can ask potential users or observe them directly (e.g., they are using an improved cookstove). Measuring adoption can be done at different scales or units of observation, including in the field plot, village, or district. Having accurate information about adoption of innovation can also be a primary interest for measuring the success of a dissemination campaign. In practice, though, measuring adoption can be complex. People may not recognize the innovation by its name. They may think they have adopted it or report that they have, but they may not be using the innovation as intended. Or an innovation may have spread so widely that it becomes

[2] In many impact evaluation settings, technology adoption results from an intervention (such as government policy). Yet adoption of new technologies can also occur without a specific intervention, by diffusion through existing markets, supply chains, or institutional networks, for instance.

[3] For example, we can build a model by using satellite-based spectral indices as independent variables and land cover classes as dependent variable at a sample of known locations. Then, this classification model will be used with spectral indices to predict the land cover classes at other (unknown) locations to produce a land cover map as the modeled output.

impractical to measure adoption using a household survey. Indeed, survey-based measures of adoption can be noisy and potentially also biased.

EO can provide an alternative by measuring adoption from space, which, in some cases, maybe more objective than other methods. To use EO to measure adoption, two key conditions must be present. First, one needs to be able to observe or detect the adoption of the innovation visually on the satellite image. Second, there needs to be a visual difference or contrast, in space or/and time, with those who have not adopted, so that there is observable spatial heterogeneity in adoption of the innovation during the study period. This would not be the case, for example, if a technology was used universally by all units of observation in all time periods or if it was never used by any of them. Ultimately, mapping adoption entails differentiating geographic areas where the innovation has been adopted from areas where adoption has not occurred. The satellite imagery needs to contain enough spectral, spatial, and/or temporal information to distinguish these two classes. If the information is visible or can be made visible through remote sensing analysis, a classification algorithm is used for distinguishing the two classes (or for comparing technology or treatment levels). Output consists of a map of the study area with at least two categories showing the areas of adoption and non-adoption. This map shows where adoption is occurring but it does not attribute adoption to an intervention – or inform about who is adopting – which means that additional information on the roll-out is required. For example, the Conservation agriculture (CA) case study assessed the adoption of the innovation by mapping cropland areas with tillage (conventional) and without tillage (under CA) in India.

Case study: Conservation agriculture in the Indo-Gangetic Plain of India

Investigators: Anil Bhargava, Camille Boudot, Andre Butler, Guillaume Chomé, Khushboo Gupta, Rupika Singh, Urs Schulthess. Project status: Completed.


Photo: Petr Kosina/ CIMMYT

The focus of this project was to map conservation agriculture (CA), an agricultural management practice that promotes minimal soil disturbance (no tillage), maintenance of a permanent organic soil cover (mulching), and the diversification of crops grown in sequence or association (preferably including at least one legume). It has been adopted in the intensive rice-wheat cropping systems of the Indo-Gangetic Plain of India. The project used remote sensing and household surveys to measure the scale of adoption in the states of Bihar, Haryana, Punjab, and Uttar Pradesh. The project team relied on optical and radar satellite images to map the area with zero tillage (ZT) and the area with conventional tillage (CT), the detectable spatial characteristic that they used to distinguish between farming types. First, they performed image segmentation of fields using Sentinel-2 to create homogenous zones. They used these segments to map cropland versus non-cropland. Within the cropland area, they relied on Sentinel-1 SAR data for characterizing tillage (ZT versus CT), using a random forest classifier.

EO does not solve the challenge of partial adoption. Partial adoption happens when only part of a bundle of innovations is adopted or when implementation does not meet the standards established for the innovation. In some cases, EO analysis can still assist with identifying a third category for partial adoption that would produce a classified map of three classes: adopter, partial adopter, and non-adopter for the unit of observations. Because the technology adoption status may change over time through new adoption (intentional rollout), dis-adoption, or spillover (diffusion), it can be useful to map adoption over time.

Measuring outcomes with EO

Measuring causal impacts is done by comparing the change in one or multiple outcome variables between baseline (before adoption) and intervention (after adoption), in both treatment (adopter) and control (non-adopter) areas. Specifically, we are interested in the change in outcome before and after an intervention, and so, in most cases, in measuring a continuous variable at two or more points in time.

From a remote sensing perspective, assessing the change in outcome variables for impact evaluation requires the use of methods for change detection (small number of images from distinct times) or time-series analysis (dense time series of imagery). The process of change detection takes advantage of the ability of remote sensing images to capture a record of conditions at different points in time, enabling researchers to detect and characterize those changes over time. The use of these methods has important requirements for rigor. Ideally, change detection procedures should rely on images captured by the same sensor (or a similar sensor) and sharing the same characteristics, including spatial resolution, view geometry, spectral bands, radiometric resolution, period of the year, and time of day.[4] The reliability of the change detection process requires the diligent pre-processing of satellite imagery to avoid introducing errors when comparing images captured at different times, including accurate spatial registration and correction for different atmospheric conditions. We discuss pre-processing further in section 5c.

Using satellite Earth observations for measuring outcomes is different from directly measuring the change from the ground using surveys or other measurement tools, because it takes place from space, far from the phenomenon under study. In general, measuring outcomes with EO data involves identifying in satellite imagery (mostly spectral) information that correlates with an outcome of interest and analyzing changes in this correlate as a means of quantifying impacts. The EO-based measurement is thus a sort of proxy variable, in the sense that it serves to replace the direct measurement of an outcome variable that would otherwise be measured from the ground. A good EO-based proxy variable must have a close correlation with the outcome variable. Satellite-based proxy variables are identified by analyzing the statistical association between direct outcome measures contained in reference data and imagery features.[5] Intuitively, some EO-based variables may be more direct proxies than others for a given outcome. For example, we might consider the Normalized Difference Vegetation Index (NDVI)[6] a direct proxy for photosynthetic activity but only an indirect proxy for crop grain yields (because grain yields are not perfectly correlated to green leave biomass). In general, direct proxies are preferred but are not available in every research context.

The sensitivity to changes in an outcome variable is an important factor for determining the adequacy of EO-based measurements for impact evaluation. To detect the change in an EO-based proxy outcome variable, the difference in values needs to be large enough to be distinguishable from underlying variability in the EO data. Challenges may arise if technology adoption has a very small effect on the direct outcome or if the magnitude of variation induced on the satellite-based proxy variable is too weak; in such cases the changes will not be detected. Before implementing an impact evaluation, it is important to assess whether changes in outcomes can be detected in satellite-based proxy variables. This can be done, for instance, by verifying at known locations of change, that this change can also be captured in satellite imagery. Another challenge concerns the time lag between adoption and detectable changes in outcomes. For example, on highly degraded common lands of India (Restoration of the commons case study), it may take many years before changes in soil carbon density can be detected. Temporal lags should be considered when assessing the suitability of remotely sensed datasets in an impact evaluation, as other methods could be more sensitive to changes.

Once outcomes and adoption are properly measured, the next step of an impact evaluation is to

[4] The period of the year is important to account for seasonal variation. The time of day is important for optical imagery.

[5] This relationship can be quantified in a variety of ways, ranging from simple correlation between the proxy and the outcome to more complex linear (regression-based) models and non-linear prediction methods that use machine learning.

[6] $NDVI = \frac{(NIR-Red)}{(NIR+Red)}$,

where *Red* and *NIR* stand for the red (visible) and near infrared wavelength regions, respectively. NDVI exploits the fact that plants absorb red (visible) wavelengths of light during photosynthesis and then re-emit them at near infrared wavelengths. The difference between the amount of NIR and red light, relative to the total amount of both, can thus be seen as a direct proxy for photosynthetic activity.

estimate the causal effect of technology adoption on outcomes of interest through statistical analysis. Estimating causal effects requires a source of exogenous variation in adoption that can be linked to differences in the outcomes of interest. Here, exogenous variation refers to variation in technology adoption that is independent from other confounding factors that may also affect the outcome. For example, technology adoption facilitated through a randomized controlled trial (RCT) is, by design, random and therefore unrelated to potential confounders. Non-experimental variation in adoption may also be plausibly exogenous. Geographic factors (such as country borders or variation in agroecological suitability regions) may influence the adoption of a particular technology but be unrelated to other factors that could also affect the outcome. Only when a sufficiently credible source of exogenous variation in technology adoption has been identified – and one that contains relevant spatial identifying information – should one proceed with an impact evaluation using EO data.

IN BRIEF:

- *Measuring adoption with EO requires the innovation to be detectable on the satellite image over the study period.*
- *Measuring impacts with EO typically implies measuring the change in an outcome with change detection or time series analysis methods.*
- *Satellite-based proxy variable for measuring change is built from a carefully validated relationship between reference data and imagery features.*
- *The feasibility of EO for impact evaluation needs to be evaluated on a case-by-case basis, and depend on the availability of spatially explicit exogenous variation in technology adoption.*

4. Advantages of remote sensing methods for impact evaluation

EO data provided by satellites have several key features that make them attractive for use in impact evaluations.

Satellites provide a comprehensive view of the Earth's surface at different scales. Indeed, the spatial coverage of remote sensing data can enable researchers to evaluate impacts over vast areas, even spanning multiple countries. In addition, it allows researchers to scale up localized measurements (e.g., GPS points) and construct wall-to-wall measures of technology adoption or outcomes at every location in an entire area of interest. Scaling up data in this way can increase studies' sample sizes, leading to more precise estimates of impacts. It facilitates the analysis of spatial leakages or spillover effects in areas near where technology adoption has occurred without major supplemental efforts.

There is a long time series of remote sensing observations. Long-term satellite missions provide data going back decades; data from the Landsat program go back 50 years (Figure 1). Moving forward, as existing missions are maintained, the temporal coverage of satellite data will continue to increase. Long-term historical data can allow researchers to assess baseline conditions and trends before the introduction of a technology. Assessing these pre-treatment conditions is an important part of the statistical analysis used in impact evaluations. Historical data from long-running satellites can also be used to evaluate the impacts of projects or interventions long after they occurred.

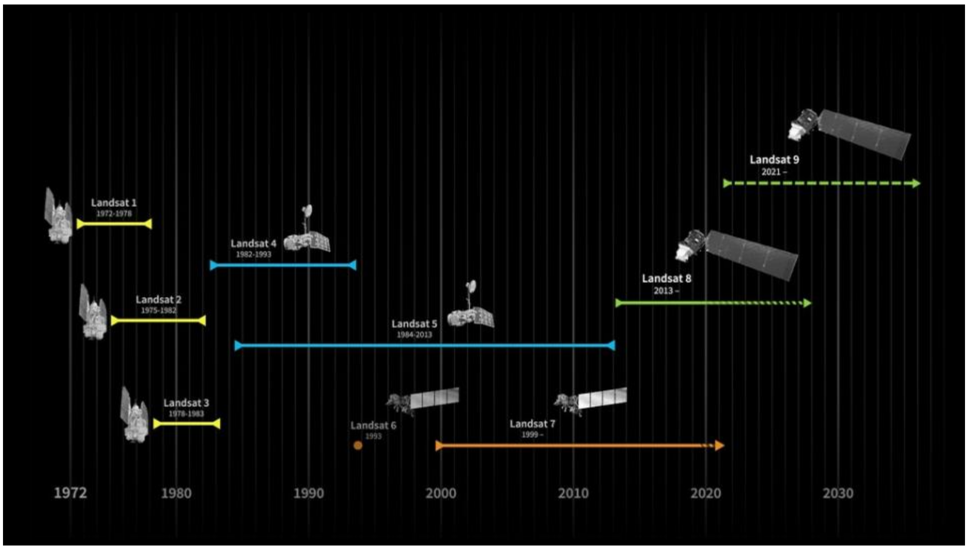


Figure 1. Timeline of the Landsat program, from Landsat 1 launched in 1972, through Landsat 9, launched in 2021. Landsat 6 failed to reach its orbit at launch. The hashed lines for Landsats 7–9 indicate the future lifespan of the satellites, which is unknown. Source: NASA's Scientific Visualization Studio.

EO data are collected regularly and repeatedly. Repeated observations can be used to build datasets where the conditions of units of observation are recorded at multiple points in time before and after adopting a particular technology. Panel data of this nature are required to implement some of the most rigorous and cutting-edge statistical techniques for estimating causal effects.

EO data provide consistent, comparable, and standardized measurements. Traditional ground-based data, such as from surveys, often require careful interpretation when the methodologies used to record information change over time or when data are collected using different methods in different locations. Remote sensing data offer a convenient way to construct measures of technology adoption or outcomes that are consistent and comparable across both time and space. This feature can be particularly convenient when attempting to look back in time and construct estimates of baseline conditions. In addition, the standardized nature of remotely sensed data means that they are not subject to certain sources of measurement error, such as enumeration errors, errors in recall, or response biases.[7] These concerns are particularly relevant in self-reported survey data when some survey respondents have received an intervention and others have not. Overall, using remotely sensed data to avoid potential measurement errors can improve the accuracy of estimates from impact evaluations. Still, building remote sensing models and interpreting their results require information about the ground conditions and/or extensive local knowledge for most applications.

Satellites provide increasingly diverse and complementary measurements. Satellite missions measure different geophysical parameters at different frequencies, spatial resolutions, and spectral characteristics. Measurements with new optical, radar, and lidar sensors are increasingly available and offer options for mapping technology adoption or outcome variables.[8] Complementary data of this nature can allow researchers to assess the robustness of impact evaluations to alternative measurement choices and/or to better understand the spatiotemporal scale at which technologies affect outcomes.

EO data are complementary to other data sources. These include ground-based data using “traditional” methods, such as surveys. EO data can be used as the main data source for impact evaluation or as a complement to other data collection methods, to increase robustness of impact analysis. Satellites may also allow researchers to obtain data from areas where it would be too

[7] For example, satellites can be used to objectively measure illegal practices that are often underreported in surveys. One example is crop residue burning in the Indo-Gangetic Plain of India. Since the practice has been made illegal in certain areas, surveys may be ill suited to produce reliable figures of the extent of residue burning practices; remote sensing can make it possible to collect objective figures.

[8] For example, the recent NASA Global Ecosystem Dynamics Investigation (GEDI) spaceborne lidar program and the upcoming Biomass satellite P-band radar will fashion new capabilities in Earth surface monitoring of ecosystem functions by capturing variables about ecosystem structure, including tree canopy height and biomass.

impractical or even dangerous to carry out field research. For example, satellite data plays an important role in detecting illegal extractive activities, such as logging in protected areas.

Free and open public access to satellite data and derived geospatial datasets has increased dramatically. Data availability is no longer considered a constraint; the challenge now resides in providing a data architecture solution to reach global users and expand the impact of satellite data. The recent proliferation of EO data platforms is making access and use of satellite data easier. Often just by creating a free account, researchers can search for and download vast amounts of data on new platforms for distributing satellite imagery.[9] Some platforms, including Google Earth Engine (GEE) (Gorelick et al., 2017) and Amazon Web Services (AWS), allow users to interact with petabytes of geospatial data and perform analysis directly in the cloud, eliminating the need for users to download and handle bulky datasets locally on their computer. Many data providers are building the spatial data infrastructures (SDIs) to facilitate the use of increasingly large volumes of satellite data. One of them is through Open Data Cube, which streamlines data distribution by lowering technical barriers for users, thus improving free and open EO data access, handling, preparation, and efficient analysis for a larger pool of global users. Annex 5 provides the name, weblink, and short description of existing data providers where researchers can access satellite data.

Finally, using EO data can help increase the transparency and replicability of impact evaluations. In geospatial science, it is becoming best practice for open science to publish the geospatial products created, the scripts used to process data, and the training/validation datasets on data repositories. Similar requirements are emerging in the social sciences, including economics and political science. Government funding agencies, including U.S. National Aeronautics and Space Administration (NASA) and the National Science Foundation (NSF), are also increasingly asking researchers to publish computer scripts and the data products from projects funded with public resources and requesting that investigators use open-source software. Some academic journals also request publication of datasets to enable other researchers to replicate study results. These trends are beneficial to practitioners of impact evaluations for two reasons. First, replication data and scripts produced by others can be used or modified for impact evaluations that use similar or related EO datasets. Second, adhering to best practices for transparent and replicable geospatial research can improve the rigor of impact evaluations that use EO data.

5. Creating successful interdisciplinary teams for geospatial impact evaluation

We reflect on elements that support successful research teams for geospatial impact evaluation, in terms of the good union of skills, experiences, and backgrounds, based on the inputs received in interviews. Teams have shared their view of what they have learned from their interdisciplinary collaborations. This input is valuable because this practical experience is rarely shared in scientific papers but remains fundamental to achieve successful projects.

The Stress-tolerant rice varieties case study team recommends including remote sensing specialists from the onset of the project, even while the proposal is being developed. This will improve the proposal, the workflow, and the timeline of the project. They found that it takes a lot of time for the team to cross the technical language barrier between economists and remote sensors/geographers, which is admittedly a major learning experience. A full-scale geospatial impact evaluation with EO data requires real interdisciplinarity, with a lot of back and forth in the workflow between team members. From a remote sensing perspective, the type of analysis whose objective is to use remote sensing for measuring very fine impacts, as required for geospatial impact evaluation, is both rare and specific. It can be hard to measure the impact of technology with remote sensing if it is not clear how the data will be used and how

[9] Many data providers have developed application programming interfaces (APIs) that respond to custom scripts for downloading imagery, or plugin tools for downloading and visualizing imagery that can be integrated into open access GIS softwares (for example, QGIS).

data limitations can play out in the context this specific analysis. Success comes from creating a two-way understanding in combining remote sensing and causal inference methods.

The Restoration of the commons case study teams notes that measuring impacts using remote sensing can be very challenging without expert knowledge. Remote sensing is technically advanced, and the field is progressing at a fast pace that requires an understanding of the different types of satellites, their use, and their methodological specificity for different purposes. It is easy for someone from outside the discipline to have unrealistic expectations about what is possible, to assume that satellite imagery can be used for all different purposes, and to believe that one can go back in time and reconstruct baseline data. The reality is that remote sensing is an entire scientific field. Having the right remote sensing expertise has been a crucial component of the project's success.

Working on the Sorghum-millet case study, the team members found that collaboration worked smoothly, as they all had previous experience with both remote sensing and causal inference methods, which makes their case unique. They see that to have an easy and productive interdisciplinary collaboration, it helps when the collaborators understand each other's epistemologies and ways of creating knowledge and research process. When starting collaborations on geospatial impact evaluation with people with different disciplinary backgrounds (e.g., economist/econometrician or remote sensing engineer), it is essential to establish a dialogue, as well as adopting an attitude of openness and willingness to learn. It is also important to include someone who has the interdisciplinary background to understand both sides and can act as a bridge to improve understanding between the different parts of the team. Otherwise, creating the knowledge from scratch may take a substantial amount of time and effort.

For the team working on the Improved forages case study, remote sensing expertise in combination with researchers from different backgrounds (agricultural and behavioral economists, soil erosion expert) and field expertise, including with long-standing field experience in Ethiopia, has been crucial to support reference data collection, analysis, and interpretation of results. Indeed, support from people with knowledge of the context and ground reality is conducive to a successful team for geospatial impact evaluation.

In other context, specialized knowledge on big data processing is a major asset. The Index-based livestock insurance case study, with the large extent of their project area and the large amount of very high-resolution imagery they have been able to acquire through their collaboration with the US Department of Agriculture, would have benefitted from a collaboration with computer scientists. The thousands of images that they processed for their project would have required higher computing capacity and capabilities than they had originally planned. So future projects that will have large spatial coverage with very high-resolution imagery should plan for this kind of support.

IN BRIEF:

- *Economists who wish to use remote sensing for impact evaluation would do well to recruit remote sensing experts at the project onset to facilitate the workflow, and build realistic expectations and timeline.*
- *Teams combining different expertise need to be willing to spend the time and efforts to create mutual understanding.*
- *Remote sensing science requires specialized knowledge that takes time to acquire.*
- *Having a person with an interdisciplinary background on the research team can improve the understanding and communication between team members.*
- *Recruiting computer scientists in big data projects is an important option to consider.*

6. Methodological considerations when using remotely sensed data for impact evaluation

While EO data may have great promise for understanding the impacts of innovations in agriculture and natural resource management, working with them can also be challenging. In this section, we discuss the main methodological challenges that can arise when using EO data and offer practical suggestions for how to address them. Throughout, we provide concrete examples from case studies and highlight good practices.

a. Using existing off-the-shelf geospatial products (R)

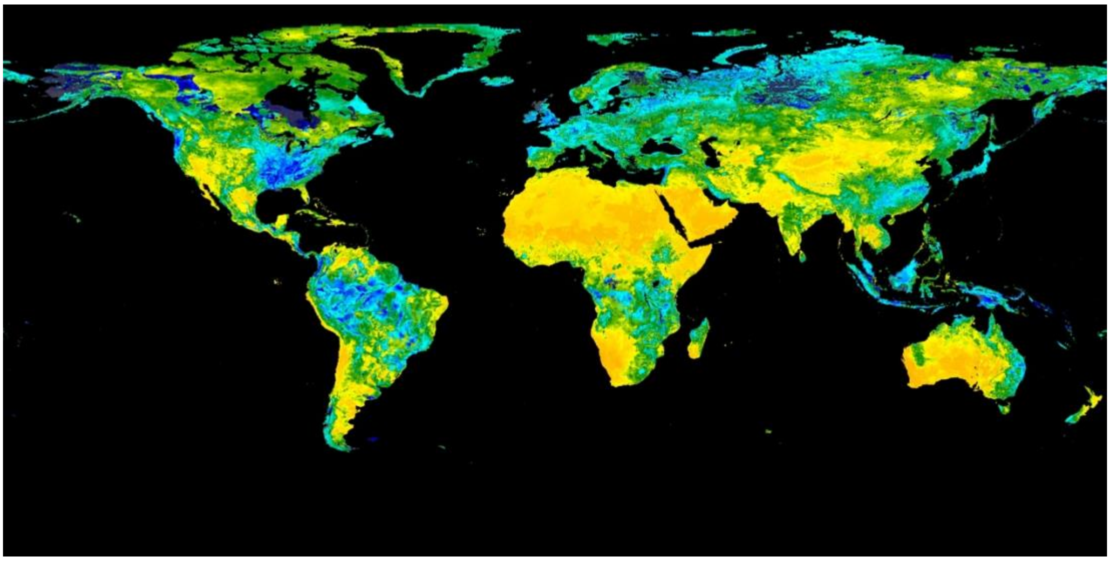


Figure 2. The Soil Moisture Active Passive (SMAP) "Surface Soil Moisture 9 km (L4, 12z Instantaneous, Model Value-Added)" layer displays model-derived global surface soil moisture of the top 5 cm of the soil column. Source: NASA Worldview snapshot, Reference: SPL4SMAU doi:10.5067/LWJ6TF5SZRG3.

Existing or off-the-shelf EO-based products are datasets created by scientific teams through the processing and analysis of satellite datasets. The emergence of these products has boomed over the past 10 years and provides new monitoring opportunities for end users by offering comprehensive and consistent satellite-based physical, biological, and socioeconomic variables at global or regional scale. These products include soil moisture (Figure 2), forest cover loss and gain, percent tree cover, biomass, rainfall, burned area, floods, and land use/land cover, to name just a few. They can be incredibly useful for many applications, including impact evaluation. For impact evaluation, it is more common to find existing geospatial products appropriate for measuring outcomes than for adoption, which is more idiosyncratic to a specific innovation.

Before using any off-the-shelf geospatial product for impact evaluation, it is imperative to have a good understanding of the product's characteristics (including the spatial, temporal resolution, the extent, and the different layers or bands) and how it was generated. This is done by reading the documentation accompanying the product (e.g., user guide or journal article), with a focus on the methods and metadata, to gain a critical understanding of the assumptions and limitations. Like any other data sets, existing EO-based products may have some intrinsic technical issues and/or be less than ideal for measuring impacts for the following reasons: (1) the characteristics of off-the-shelf products may not apply to the context of the intervention under evaluation; (2) existing products may not well represent the local context; and (3) several competing products may exist with divergent values. The first point is generally obvious when examining the documentation. For example, the products may not have an adequate spatial resolution or may not cover the period of interest. It remains the responsibility of researchers to justify the appropriateness of using an existing EO-based dataset and to demonstrate the validity of their choice for their application.

When the products' characteristics are applicable to the impact evaluation case, it is important to assess how accurately the off-the-shelf product captures the local conditions in the study area. Many existing global or regional EO-based data sets are not sufficiently validated at local and national levels. For projects working at the scale of a watershed or protected area, for example, there can be large discrepancies between the values of global products and what is observed on the ground.[10] The quality and validity of an existing product are assessed by comparing its values with reference data for a sample of locations and/or of other data sources.[11] The Stress-tolerant rice varieties case study in Bangladesh found issues in an existing product after the product was evaluated.

Sometimes, different geospatial products exist for the same variable. When different products exist, it is important for researchers to demonstrate the robustness of their impact evaluation results to the choice of remote sensing products. For example, Michler et al. (2021) showed that one can obtain both positive and negative estimates of the effect of rainfall on agricultural productivity simply by using different EO-based rainfall products in the analysis. Robustness can be demonstrated by (1) showing that the choice of EO product is the most accurate compared with ground reference data if one product is selected over others; (2) comparing the EO products between themselves to identify possible bias and spatial uncertainty (locations where values differ more); (3) using sensitivity analysis to compare impact results using the different EO products; and/or (4) employing an aggregate measure of the existing EO products, using basic statistics (e.g., mean) or more complex techniques (Zhang and Liang, 2020). When reference data is available for comparison, it should be prioritized for evaluating EO product's validity, but the other approaches described above are also important to demonstrate robustness.

If no adequate EO products exist, creating custom ones for impact evaluation requires much higher level of expertise in remote sensing analysis than using off-the-shelf products. Evaluation team should make sure they have the appropriate level of skill and sophistication before moving forward.

IN BRIEF:

- *Obtaining a good understanding of the methods and assumptions used for generating the EO-based products is crucial.*
- *Comparing geospatial products with reference data for a sample of locations and/or with other data sources are an important step to assess quality/validity of the product.*
- *When different appropriate EO products exist, it is crucial to demonstrate the robustness of impact evaluation results to the choice of remote sensing products.*
- *The cost-benefit of adapting existing products compared to creating a new one depends on resources, expertise, and interest.*

[10] Several studies have shown large differences between estimates of an off-the-shelf global forest loss product (Hansen et al., 2013) and local or national land-cover change maps or field data (Tropek et al., 2014; Burivalova et al., 2015; Cunningham et al., 2019; Pelletier et al., 2019). Discrepancies can arise from (1) the lack of ground calibration and validation field data (especially from developing countries) to build and evaluate the models of these existing products; (2) the diversity in tropical ecosystems and social-ecological systems, and (3) differences in definition (e.g., the definition of forest varies between countries) as well as other product-specific factors.

[11] Approaches to validate and correct biases have also been proposed to adapt global products to national and local contexts (see, for example, McRoberts et al., 2016; Sannier et al., 2016). In general, these approaches require reference data and additional contextual information. It should also be noted that for some variables of interest, locally calibrated maps may not perform better than global products (e.g., Burivalova et al., 2015).

Case study: Stress-tolerant rice varieties in Bangladesh

Investigators: Jeffrey D. Michler, Mathieu Renaud, Valerien O. Pede, Anna Josephson, Tom Evans. Project status: In progress.

Off-the-shelf global remote sensing products may not be sufficient for use in impact evaluation. Indeed, this project assesses whether Stress-tolerant rice varieties (STRV), introduced more than a decade ago, reduce yield loss due to flooding on rice fields in Bangladesh. To estimate impacts, the research team compared post-flood vegetation greening on fields where STRV were adopted versus fields where they were not. In phase 1, they computed historical flood incidence and intensity index over rice areas using the Dartmouth Flood Observatory (DFO) flood area products archive, which goes back to 1985. However, by comparing the DFO product to Sentinel-1 flood maps for more recent years, the team noticed that the DFO product underestimated flooded area. They also learned that in the earlier years (before 2010), the DFO was less comprehensive. Specifically, the DFO website states the following warning: "The statistics presented […] are derived from a wide variety of news and governmental sources. The quality and quantity of information available about a particular flood is not always in proportion to its actual magnitude, and the intensity of news coverage varies from nation to nation." The potentially biased representation of historical floods was an important consideration for the STRV project. This example illustrates the importance of looking for information about how products are generated and verifying data with alternative sources. For phase 2, the STRV project will use the recently produced Global Flood Database, based on analysis of MODIS images (Tellman et al., 2021).

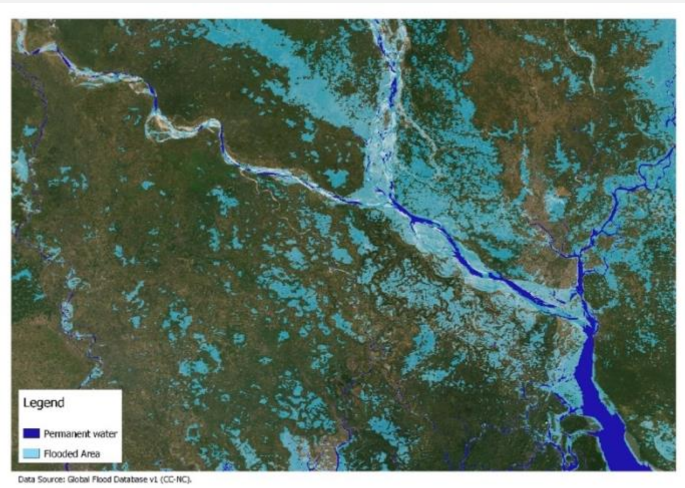


Figure 3. Flood area in Bangladesh, from June 20 to October 7, 2004. (Author: J. Pelletier; Data source: Global Flood Database v1.)

b. EO data availability and trade-offs

With all the fanfare surrounding remote sensing data, it is possible to have inflated expectations about availability and potential uses. Researchers that consider creating custom EO products for impact evaluation should (1) verify the availability (and sufficient quality) of remote sensing data for the area of interest; and (2) evaluate the trade-offs between different sensors to meet the impact evaluation goals.

Relevant EO data may not be available. While the quantity and accessibility of remotely sensed data has increased rapidly in recent years, EO data for many study areas and/or periods of time may be missing. Annual cloud cover, seasonality, and topography are notable constraints to optical image availability and quality for historical impact studies (Herold, 2009; GOFC-GOLD, 2016; Prudente et al., 2020). This is most important in the tropics during the rainy season, when cloud cover severely limits the availability of satellite imagery. Image availability and quality may also be more limited in older satellite imagery. For example, some regions of the world were not well covered by Landsat ground receiving stations in the 1990s, resulting in only few images being captured for the period for these areas. In general, constraints on data quality and availability are one of the main limitations for retrospective impact evaluation using EO data.

Retrospective impact evaluations can face important limitations because they have limited options in terms of imagery. Some outcomes, including for example tree cover and staple crop productivity over relatively large intervention units, can plausibly be estimated with 30-m Landsat images with a revisit time of 16 days, if cloud-free images are available. Other outcomes (e.g. fishponds) can only be effectively measured with Sentinel data, starting the time series after 2015-16. Several projects reviewed for these guidelines study the impacts of interventions that started 10 years ago or more. The Direct seed marketing (DSM) program in Ethiopia, which began in 2011, is one example. The evaluation would have greatly benefited from 20 years of Sentinel-2 data (10m VNIR), but the satellite was launched only in 2015, so the project has to rely on MODIS (250m resolution) and Landsat (30m resolution) data to assess both baseline (pre-treatment) and post-intervention conditions. In general, retrospective impact evaluations are likely to face trade-offs of this kind, or be outright unrealizable with existing EO historical capabilities.

Satellite data exhibit inherent trade-offs between spatial, temporal, and spectral resolutions, even when data are available for the area and time period of interest. Some sensors offer historical data over a long period but only at a coarse resolution. Others provide high spatial resolution but with a cadence of a few weeks for the same location. These trade-offs arise because satellite missions are launched to achieve a set of specific data collection objectives. When assessing EO data for use in impact evaluation, researchers need to think carefully about the nature of the processes under study (e.g., those related to technology adoption or outcomes) and how they relate to imagery characteristics. For instance, someone who studies weather events may be willing to sacrifice spatial resolution to obtain higher temporal resolution, whereas someone studying seasonal vegetation changes would likely value higher spatial and spectral resolution but be overwhelmed by hourly imagery data. To help with these decisions, we provide information on satellite missions and their associated imagery characteristics in Annex 7.

Finally, the cost of very high-resolution imagery may make the most suitable options for measuring adoption or outcomes in a specific project unattainable. In some cases, very high-resolution imagery (VHR) necessary for detection may not be available for the period or area of interest. But even when such data do exist, getting access to it could be cost prohibitive. In fact, even though the supply of very high-resolution imagery has increased greatly in recent years, the cost of imagery has not necessarily gone down. However, there is hope that costs will decrease as new public satellites are launched (e.g., Sentinel HR) and more agreements are signed with commercial providers to make existing data available for free. The high cost of VHR imagery was one challenge faced by the Demi-lunes case study.

IN BRIEF:

- *Satellite imagery may have insufficient quality or availability for the area of interest and evaluation goals and should be verified before the project onset (when writing the proposal).*
- *Inherent tradeoffs between different satellite products may require leveraging multiple sensors and data products to meet project goals.*
- *The cost of very high-resolution imagery can be prohibitive to access.*

Case study: Demi-lunes rainwater harvesting in Niger

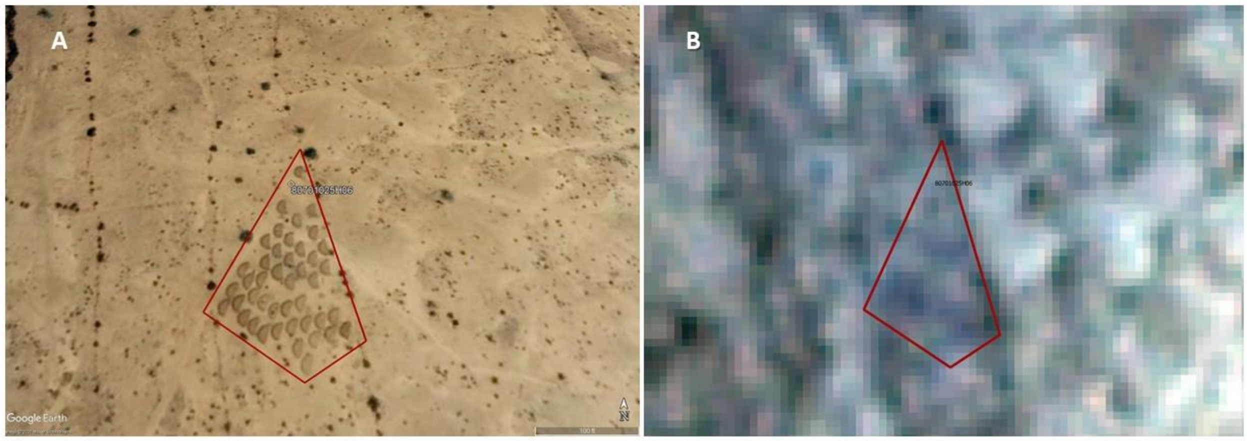


Figure 4. Comparison of the same location (hhid=80701025H06), on the same date (August 11, 2018) for a treated site with demi-lunes water-saving technology. The demi-lunes are clearly visible on the Google Earth image, with a ~ 0.5 m-pixel resolution (A), but not on the PlanetScope image (~3m resolution) (B; true color composite image). Source: J. Pelletier.

Investigators: Jenny Aker, Kelsey Jack, Kendra Walker. Project status: In progress.

The demi-lunes project studied the impacts of semi-circular rainwater harvesting structures (1m by 4m in size) in the Zinder region of Niger. The technology is visible in Google Earth, which has very high-resolution imagery with ~0.5-m resolution (Figure 4A) but not on PlanetScope data, which are free of charge for academic purposes (~3m resolution) (Figure 4B). Initially, the goal of the study was to map the adoption and dis-adoption of demi-lunes at a regional scale and calculate the spillovers associated with their use. Ultimately, however, the 0.5 m-resolution imagery was deemed too expensive given that the aim was for the mapping method to be transferred to the Niger government for monitoring the practice at scale. So the research team shifted to measuring the impacts of demi-lune adoption on soil moisture and productivity, using survey data in combination with NDVI and soil moisture indices derived from PlanetScope and Sentinel-2 imagery.

c. Pre-processing (R)

Image pre-processing refers to operations that correct distortions and noise in satellite images to create a more faithful representation of the Earth's surface and improve image quality and utility. All data collected from instruments on board satellites need to go through some modifications before they can be used by most researchers, and additional corrections are required for rigorous analysis. Practitioners of impact evaluation need to ensure that the imagery they use has been duly pre-processed or that they pre-process it before their analysis. Image pre-processing is especially important for remote sensing analyses that involve study areas larger than one satellite image or that require comparing multiple images over time, since these corrections are necessary to make images comparable over time and space. Because impact evaluation entails a comparison in time of outcome variables (change detection, time series), pre-processing steps are mandatory. Failing to perform these corrections up front will introduce systematic errors in impact results down the road.

To acquire EO data that is already pre-processed, researchers can look for analysis-ready data (ARD), which is ready to use for analysis. The Committee on Earth Observation Satellites (CEOS) defines ARD as "satellite data that have been processed to a minimum set of requirements and organized into a form that allows immediate analysis with a minimum of additional user effort and interoperability both through time and with other datasets."[12] There are major ongoing efforts by satellite data providers, government space agencies, and international organizations to produce ARD to reduce the burden of pre-processing on global remote sensing users. Examples of ARD include the Landsat 8,9, and Sentinel-2 harmonized products, Open Data Cube, and Landsat ARD for the continental United States (Dwyer et al., 2018) (Figure 5), and the Global Land Analysis and Discovery (GLAD) Landsat ARD data.

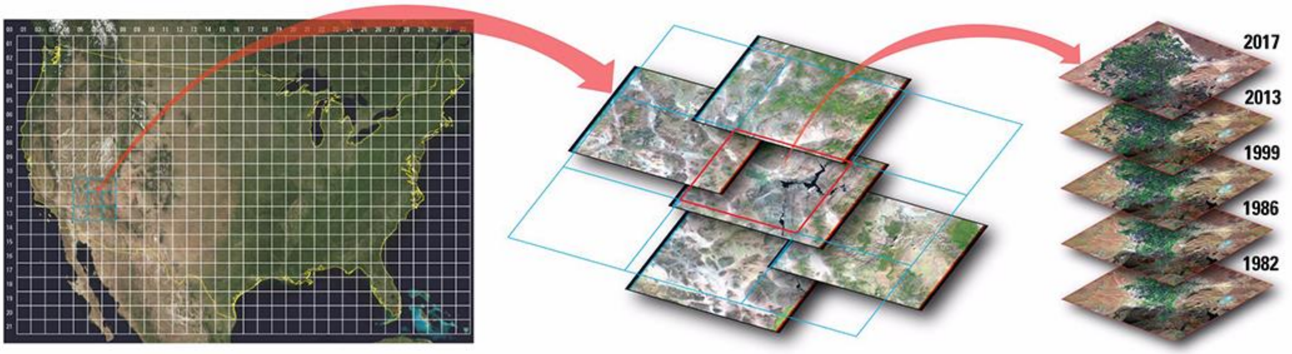


Figure 5. Illustration of Landsat analysis-ready data products for the continental United States. Global Landsat ARD is still under development at the time of writing. Source: U.S. Geological Survey Communications and Publishing.

There are two main types of sensors: passive and active. Passive sensors are those that rely on the sun's energy that is radiated or reflected by the observed surfaces, as for optical imagery. Active sensors, in contrast, provide their own source of electromagnetic radiation to illuminate the terrain, and include radar imagery.

For optical imagery, pre-processing includes radiometric, atmospheric, and geometric corrections. Radiometric correction converts the raw output recorded by an instrument on board a satellite into what is known as a top-of-the-atmosphere (TOA) reflectance (a view of the Earth's surface as if from above the atmosphere). It does that by converting raw output into radiance values using instrument- and band-specific calibration factors (radiometric rescaling coefficients) and for differences in scene illumination that result from variation in the time of day and relative positions of the satellite, Earth, and sun when images were taken.[13] Atmospheric correction is carried out to compensate for atmospheric effects to convert the TOA reflectance image to bottom-of-the-atmosphere (BOA) or surface reflectance image. It corrects for the scattering and absorption effects of gases and aerosols present in the atmosphere.[14] Geometric correction is applied to remove the significant geometric distortions in a raw digital image. It uses ground control points (GCPs) and a digital elevation model (DEM) to anchor the image into a spatial reference system.[15] Geometric correction is especially important in areas with relief and for accurately matching the specific locations on Earth and on the image, as it is the case with impact evaluation. Fortunately, geometric correction performed by data providers has greatly improved in the past 10 years. Still, in some cases, pre-processing can represent a significant hurdle for analysis, as in the Index-based livestock insurance case study.

---

[12] https://ceos.org/ard/

[13] For example, corrections are made for the angle between the satellite, location of imaging, and sun (sun angle), the elevation of the sun, and the seasonally varying Earth-sun distance at the time of imaging (Lillesand et al., 2015).

[14] Several methods for atmospheric correction have been developed, including the dark-object subtraction method and more complex atmospheric radiative transfer models (e.g., S6) (Kotchenova et al., 2006; Kotchenova and Vermote, 2007) or Land Surface Reflectance Code (LaSRC) (Vermote et al., 2016), with specialized algorithms for some imagery (e.g., Sen2Cor for Sentinel-2).

[15] GCPs are locations collected on the ground by GPS or from another reference image, that can be precisely identified on the satellite image. Using a digital elevation model (DEM) to account for topography, models are used to relate the GCPs, and their X and Y locations on the image using resampling methods to make sure locations on the image match locations in the spatial reference system. The model's "goodness of fit" and remaining locational error is quantified with root mean square error (RMSE) statistics.

Another important pre-processing step for change detection procedures includes the bidirectional reflectance distribution function (BRDF) normalization. Intuitively, we know that objects look differently when viewed from different angles and when illuminated from different directions. The solar and view angles for the same ground target change with the date, vary between sensors, and are influenced by the target's surface roughness and the wavelength that illuminates it. BRDF is a mathematical description of how reflectance varies for all combinations of illumination and viewing angles at a given wavelength, which normalizes the angle effect that causes discrepancies in measurements.

Spatial co-registration is a process of co-registering a temporal stack of images to the same reference image by resampling so that the same location can be compared in time. This pre-processing step is applicable to different types of imagery, including optical and radar.

Other types of imagery require different pre-processing steps. For instance, for radar imagery like Sentinel-1 C-band SAR, a pre-processing workflow includes applying orbit file, thermal noise removal, border noise removal, radiometric calibration, speckle filtering, range Doppler terrain correction, and conversion to decibels (dB) (Filipponi, 2019).

In addition to these standard corrections, some images may have specific defects (striping, missing lines)—for example, resulting from temporary technical failures with onboard instruments[16]—or may be affected by other interferences between the sensor and the Earth's surface, including clouds, shade, smoke, and haze. These defects can cause information loss, generate errors, or make images unusable.

Case study: Index-based livestock insurance project in Ethiopia and Kenya

Investigators: Steve Wilcox, Gerardo Soto, Nathaniel Jensen, Francesco Fava, Chris Barrett, Ying Sun.
Project status: In progress.

Image pre-processing has posed issues for a project on index-based livestock insurance (IBLI). The case study seeks to evaluate the environmental spillover of IBLI, an instrument designed to help protect pastoralist herders against drought-related mortality of livestock, on long-term rangeland ecosystem health outcomes in Ethiopia and Kenya. The project is constructing a new EO-based composite measure of rangeland health by combining several remote sensing indicators. Unfortunately, the Covid-19 pandemic travel restrictions upended the project's field data collection. To adapt to these restrictions, the team relied on previously collected geolocated photographs to observe ground conditions, and, through their collaboration with the U.S. Department of Agriculture, they acquired large amounts of VHR imagery (Geoeye-1, WorldView I, II, III) that they used for classification. Yet the large spatial coverage of the project (about the size of Zambia or Texas) requires mosaicking thousands of these images, each of which may be 3–10 gigabits. These images, which were collected from different sensors with different spatial resolutions, viewing angles, and illumination conditions, show large differences in reflectance even within a small time period (no real change is happening), posing a real challenge for classification.


Photo credit: ILRI/Kabir Dhanji


Borana woman herding small ruminant, Ethiopia.
Photo credit: ILRI/ Camille Hanotte.

[16] This is the case, for example, with a subset of the Landsat 7 imagery collected after May 31, 2003 (known as Landsat 7 ETM+ SLC-off), when the onboard Scan Line Corrector (SLC) failed. These images have data gaps that result in approximately 22% of a scene being unusable. Each scene needs to be corrected with noise removal.

Clouds in particular have been a persistent challenge for optical imagery (Figure 6). While there are no perfect solutions, there are several potential ways to deal with clouds and cloud-related issues (e.g., the shade they project on the land surface). First, researchers can filter images by their percentage cloud cover and use only cloud-free imagery in their analysis. Alternatively, if clouds affect only one part of an image, masking techniques can be used to convert the pixels in cloud-covered sections of an image to no-data values. Other approaches are to fill the section with data from a cloud-free image on a nearby date, or to take many images from the same area and create a composite or select the best pixel after corrections are applied. For some imagery products, users have developed custom algorithms to correct for clouds, shade, and haze.[17] The best approach depends on the data used and the type of analysis performed.

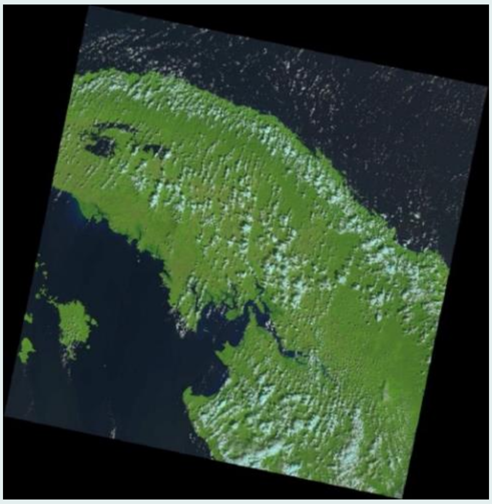


Figure 6. Landsat 8 OLI scene of the Darien Gap, one of the cloudiest areas of the Earth, located at the border between Panama and Colombia, on 2021/12/11, path 11/ row 54, with land cloud cover of 26.4%. Source: USGS EarthExplorer.

Images are made available to the users under different pre-processing levels. Some of the processing is done by the image provider. For example, all of NASA's Earth observation satellite data is pre-processed at least to level 1 standard data products, and most to a level 2 (derived geophysical variables) and 3 (variables mapped on uniform space-time grid scales) products; the raw data (level 0) is not usable for most users. It is important that practitioners of impact evaluations understand the pre-processing level of the images they access in order to identify which pre-processing steps remains to be conducted.

Once the pre-processing steps are completed or when acquiring ARD, Peter and Messina (2019) recommend using simple statistics (standard z-scores, modified z-scores, Tukey's outliers, and Geary's C) to detect and visualize spatial outliers that would need to be masked out or corrected before analysis.

IN BRIEF:

- *Pre-processing of satellite data is mandatory for evaluating impacts of innovations, and presenting the methodological description of these steps is part of a rigorous analysis.*
- *Analysis Ready Data is increasingly available and can relieve a burden for the analyst.*
- *Still, remote sensing data users should know the processing level of the imagery they are using and verify for possible distortions and artifacts by performing preanalytical verification.*
- *Failure to adequately perform these corrections on remote sensing data may lead to erroneous results and incorrect conclusions can be drawn.*

d. Reference data

Using satellite-based Earth observation data for impact evaluation almost always requires the use of reference data. Reference data serve three main general purposes (Sabins and Ellis, 2020):

1. Interpretation: EO data support interpretation between what is observed from a distance on the satellite image and the reality on the ground, assuring that the pattern one sees on the image is truly present on the ground and not just an artifact in the image.
2. Calibration or training data: EO data build a statistical relationship (model, algorithm) between what appears on the satellite image and the ground conditions, which are then used for prediction elsewhere.

[17] For example, Jin et al. (2019) compared approaches to improve cloud masking for Sentinel-2 imagery from a locally trained decision tree model, a Hollstein et al. (2016) decision tree model, and a default quality assurance (QA) band provided for removing contaminated pixels with clouds, haze, and shadows. They found the best results using the locally trained model.

3. Validation: EO data serve as validation data to evaluate the model performance—that is, to see if the model's predictions match the conditions observed in the reference data.

Reference data are often referred as ground truth. The term is not meant literally; many types of reference data are not collected on the ground but approximate true ground conditions. We can separate two general sources of reference data: (1) other remote sensing data, including very high-resolution (VHR) satellite images, aerial photos, or drone imaging; and (2) georeferenced ground measurements, which are locations collected in the field, with global positioning system (GPS) or related devices that can be associated with survey responses and/or other ground biophysical measurements. The type and source of reference data depend on availability and the type of output being created.

VHR images available through Google Earth Pro, for example, are frequently used as a source of reference data. Their use helps identify features at the Earth's surface with more certainty via image interpretation. For classification, for instance, creating reference data involves human interpreters using the VHR images to identify a set of visible classes, collecting samples of pixels that represent each class by drawing a point or polygon, and assigning each sample to a class. Reference data can be collected from VHR images for adoption or outcome variables only when these are observable or recognizable on the VHR image (e.g., percent tree cover). Many other adoption or outcome variables are not interpretable on VHR remote sensing products (e.g., most crop types and all crop varieties) and require direct field measurements.

For ground-based reference data, basic information is obtained by collecting the spatial coordinates of sample locations in the field, along with other associated information (Figure 7). The associated information varies from simple (e.g., geotagged photos) to complex measurements requiring specialized equipment. For example, the Alternate wetting and drying case study conducted water-level depth measurements in rice fields over one cropping season as a source of calibration data.

Researchers interviewed for these guidelines value the collection of reference data obtained from the "ground" through georeferenced field measurements. They perceive products created without ground verification are less reliable.

Figure 7. GPS coordinates are collected in the field to capture the area of impact of charcoal production (left). Track data functionality allowed for automatic collection of GPS points at a specific time interval (e.g., every 30 seconds) as the field technician walked around the perimeter of active charcoal production sites. The information is later uploaded in a GIS to draw the polygon of each impacted area, as shown by the red lines (right). Source: J. Pelletier.

Case study: Alternate wetting and drying in Vietnam

Investigator: Robin Lovell. Project status: Completed.

This study aimed to measure the adoption of the water-saving technique in rice fields known as alternate wetting and drying (AWD) in the Mekong River Delta, Vietnam. The project relies on a time series of wetness index derived from Sentinel-1a and 1b synthetic aperture radar data to detect changes in flooding and drying patterns throughout a rice-growing season. The reference data used to calibrate the method consisted of water-level measurements by dual-pronged water meters installed in the soil of AWD fields (where water levels varied between wet and dry) and conventional fields (which were continuously flooded) (Figure 8). Setting up these measurements in farmers' fields and covering different soil types was one of the challenging aspects of the work.

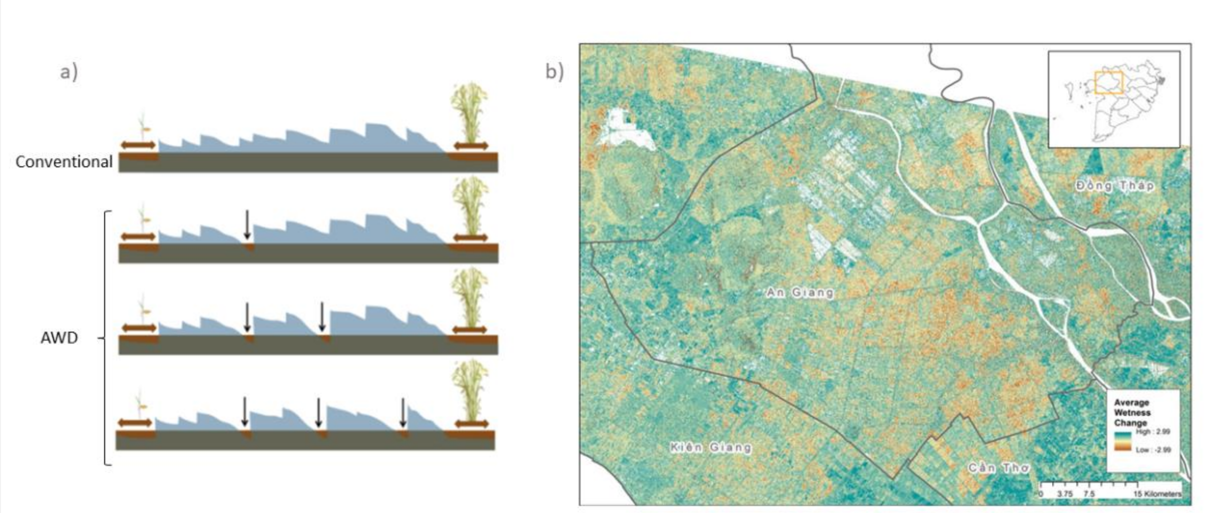


Figure 8. Detecting adoption of AWD with remote sensing involves measuring the variation in water levels over the growing season. Graphical illustration of conventional rice paddy (top drawing) and AWD rice paddy (three bottom drawings) (a). Source: IRRI GHG questionnaire. The remote-sensing method captures the variation in water levels in paddy rice (b). Source: Lovell (2019).

Guidance on reference data acquisition

Access to an adequate quantity of high-quality reference data is critical for remote sensing analysis (Foody, 2009, 2021). Both the sample size of reference data and errors in reference data can have large impacts on the reliability of the results. Reference data acquisition should be based on a sampling design that refers to the protocol that guides the reference data collection. This protocol should include sample size and an optimal allocation of reference data to be collected—that is, how much and where—to achieve a certain level of accuracy. The sampling design should ensure that the reference data collected are representative of the geography and of the variability of the outcome or feature of interest in space and/or time.

For validation, also called accuracy assessment, it is critical to use a probability sampling design, which uses randomization to select the sample (e.g., stratified random by classes/strata, simple random, systematic design).[18] For training or calibration, the sampling design is somewhat flexible and can include purposive/directed sampling.

Reference data are statistically considered the "true value," but in reality, they may also contain measurement errors that can affect model output. Specific attention to assuring a good quality and quantity of reference data can improve models and outputs, as well as general reliability. Errors in

[18] Decisions about sample size and optimal allocation should be guided by calculations performed in the planning phase and reflect the choice of sampling design, similar to any other sampling design (Cochran, 2007).

training data can have large consequences for the results (Elmes et al., 2020), and even small errors in the validation reference data set can lead to large biases on performance metrics and overall reliability (Olofsson et al., 2014). Errors in reference data can stem from two main sources: (1) errors in the design of reference data collection (decisions on how to collect), and (2) errors generated during the reference data collection process itself. We explain some measures that practitioners can take to minimize and account better for errors in the reference data.

Reference data measurements should capture whether the variable of interest is time-sensitive—that is, where the ground conditions change with time—or stable over time. For time-sensitive variables, the reference data should be collected within a window of time that matches the imagery used to create the map and the characteristic rate of change of the feature being mapped (Elmes et al., 2020). Temporal unrepresentativeness, which is a mismatch between the time when the reference data was collected and the period of interest for mapping, is a potential source of error related to the design. Knowing when the reference data were collected and ensuring valid temporal match are important to prevent this type of error.

There are additional considerations for sampling reference data for impact evaluation. A reference data sample need to have a balanced design, so that the sample is representative of the areas where adoption took place (treatment areas) and where it did not (control areas). The purpose is to obtain modeled outputs with similar levels of accuracy between treated and non-treated areas. Failure to represent both groups with similar accuracy levels can cause problem when using statistical techniques to infer the causal effects of technology adoption.[19]

When collecting reference data from other remote sensing imagery, it is important that the imagery used as the source of reference data provides higher confidence about ground conditions via higher spatial resolution, than the data used for creating the map. For classification, the protocol for creating reference data should specify how to ensure consistency among different interpreters in order to reduce interpretation error[20] and how to account for uncertainty in assigning the reference classification. For example, the protocol can include a measure of confidence for each reference label (e.g., low, medium, high confidence) and specify how to define agreement between the map and the reference data classifications.

When VHR data are used as a source of reference data, it is important to geometrically adjust images used as a source of reference data and those used to make the map so that equivalent geographic points coincide. Spatial co-registration error happens when there is spatial misalignment between these two image sources.

When georeferenced ground measurements are collected for reference data, having accurate data on location (positional accuracy) is fundamental. For this reason, using a handheld GPS device with known accuracy is preferable. For some applications, high positional accuracy is more important than the sample size. For crop mapping, for example, it has been shown that using the plot mean of full high-quality plot boundaries (highly accurate location) from a smaller number of plots (4,000 plots) leads to more accurate mapping results compared with collecting GPS coordinates of the plot corners (less accurate location) from a larger number of plots (7,000 plots) (Azzari et al., 2021).

For georeferenced continuous variables used to calibrate satellite observations by scaling field measurements (e.g., water-level meters, soil sample), it is a good idea to adopt a sampling design that reflects the spatial variability of the variable in relation to the pixel size of the imagery used for mapping. For example, a cluster of subplots that covers at least the size of one pixel on the image can better reflect ground conditions for the area than in situ measurement at one spot. Error in scaling field measurements is a co-registration issue that happens when there is a mismatch between the resolution of the imagery and the spatial extent at which in situ measurements are collected (Baccini et al., 2007).

---

[19] Remote-sensing experts will often be inclined to oversample or to focus reference data collection on treated areas, which are often rare in the landscape, in order to make sure that the treated areas are well represented and highly accurate.

[20] McRoberts et al. (2018) recommends that at least three experienced interpreters be employed and that rigorous estimation of uncertainties include an estimation of interpreter error by comparing class attribution between these interpreters for the same locations.

When collecting reference data consisting of continuous variables in the field, measurement error can stem from the miscalibration of devices, in situ variability, defects in hardware, and more. To capture these sources of uncertainty, repeated measures of continuous variables should be used to estimate measurement errors in reference data. These issues are present in all primary data collection of measurements that are used directly as outcomes (without EO), but with EO, the measurement error may translate into the image processing results.

Understanding the sources of error is important for reducing them and for producing a realistic accounting of uncertainty. The gold standard is to publish the reference data in an open access repository to allow the replication of results. At a minimum, the reference data creation methods and details on sources of error should be documented for transparency.

Barriers to reference data acquisition

Reference data are often time consuming and costly to collect correctly. In the case study projects reviewed, researchers identified access to adequate reference data as one of their major challenges.[21] The reasons for this deficiency are diverse: lack of historical reference data, inaccessibility due to bureaucratic processes or data ownership, constraints due to the pandemic or other crises, deficiencies in planning of field data collection, or a combination of factors. This situation highlights both the importance of reference data for impact evaluation projects and the challenge of accessing or acquiring high-quality reference data for remote sensing analysis.

Sometimes the problem is not the absence of data but a lack of data access. Access to geolocated data from existing surveys may be unobtainable if data owners do not wish to share it, either publicly or privately. In other instances, survey data stewards, including national statistical and census offices, apply a random offset to the GPS coordinates in order to maintain respondents' confidentiality (spatial anonymization). While the protections are essential, the resulting error in position can be a problem for remote sensing analysis when research teams attempt to repurpose these surveys as geocoded reference data (Box 1), and for some applications, the data may be unusable.

IN BRIEF:

- *Reference data play an essential role in remote sensing analysis for training and validation and come from two main sources: 1) Other remotely sensed imagery of higher resolution, and 2) Georeferenced field measurements.*
- *Error in reference data can arise from issues of spatial co-registration, temporal unrepresentativeness, interpretation error, and scaling, among other. These should be reduced as much as possible during data collection, whereas the remaining error should be estimated.*
- *Methods for the creation of reference data and associated errors should be documented for transparency.*
- *Access to adequate reference data in terms of quality and quantity, is one of the most important challenge for geospatial impact evaluation.*
- *Reference data collection is expensive and time consuming, but essential to obtain satisfactory model performance and results*

[21] The lack of reliable reference data is also perceived as the main constraint to improved EO-based model performance for measuring impacts on sustainable development outcomes (Burke et al., 2021).

Box 1. Adequate geospatial data and privacy concerns: Where is the right balance?

Spatial anonymization is a process used to preserve the confidentiality of survey respondents location when providing public access to GPS coordinate data that will enable users to integrate other spatial datasets, including EO data (Blankespoor et al., 2021). Spatial anonymization is performed with geomasking methods that alter or blur the coordinates as a way to conceal exact locations. Spatial anonymization protocols are used by large-scale household surveys including the USAID Demographic and Health Survey and the World Bank Living Standards Measurement Survey—Integrated Surveys on Agriculture (LSMS-ISA). By randomly shifting the location, spatial anonymization helps protect participant confidentiality, but can also limit the kinds of applications that the location data can be used for. Generally, the use of anonymized locations poses a data utility trade-off or analytical cost by introducing uncertainty to modeled relationships.


Photo: J. Pelletier

Recently, the expanding EO landscape and potential applications has spurred demand for access to more precise and exact location information for use as reference data in remote sensing applications for sustainable development (Burke et al., 2021). Grace et al. (2019) have looked at alternative methods to implement spatial anonymization that can better represent environmental conditions around settlements. Michler et al. (2022) showed that geomasking may have no or limited impacts on estimates of the relationship between weather and agricultural productivity. Satellite-based weather data have coarse resolution, and so spatial anonymization introduced an acceptable level of measurement error. For other applications, including crop type mapping, for example, accurate locational information is a crucial requirement and a major constraint to improved EO-based model development, with large consequences for results (Azzari et al., 2021). There is a clear need for more research to understand the implications of spatial anonymization on data utility for different applications. Further considerations are required to define options and best practices that can mitigate privacy concerns and expand the value of these household survey georeferenced data to other applications.

e. Validation of remote sensing products (R)

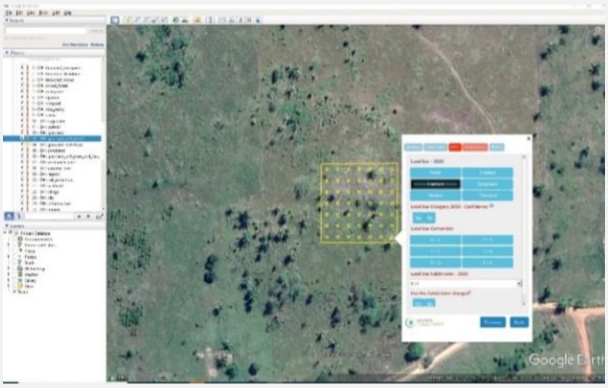
One key step in remote sensing analysis is to validate the results, which largely means evaluating the accuracy of the product compared with reference data through a procedure called accuracy assessment. In the context of geospatial impact evaluation, validation refers to the analysis of how closely technology-adoption or outcome measures produced by remote sensing data match the validation reference data.

A best practice in the field of remote sensing is to evaluate the accuracy of model outputs using independent validation datasets, different from the one used to train or develop the algorithm (Olofsson et al., 2014; Jain, 2020). This validation dataset can be a subsample of the reference data that was set aside for this purpose—and not used for model training—or it can come from other data sources, as in the Sorghum-millet intensification case study.

Case study: Sorghum-millet intensification project in Mali

Investigators: Dilini Abeygunawardane, Patrick Meyfroidt, Robert Heilmayr, John Nzungize, Daniel Müller, Adia Bey. Project status: In progress.

Sikasso, Mali. Source: Google Earth Pro.



The sorghum-millet intensification project in Mali makes use of VHR imagery and ground measurements as a source of reference data. The study aims to provide quantitative evidence of the impact of a sorghum and millet upscaling program in Sikasso, Mali. The study team investigates whether such technology-induced agricultural intensification impacts land-use/land-cover change at the village level using three measures constructed using EO data: tree cover, cropland extent, and landscape configuration. They developed a desk-based protocol for interpreting VHR imagery made available via FAO Collect Earth, an open-source application that supports visual interpretation for land monitoring. It facilitates access to multiple freely available repositories of satellite imagery (Figure 9). This approach allows the project to collect thousands of reference training data points (or polygons) using visual interpretation to support a machine-learning algorithm to analyze change in percent forest cover both outside and inside agricultural lands. They will use the field data collected in Mali for validation purposes only, keeping the ground data completely independent from the training data collected with Collect Earth.

Figure 9. Example of the FAO Collect Earth application (https://openforis.org/tools/collect-earth/) that enables data collection through Google Earth for land monitoring. Users can access data from Google Earth, Bing Maps, and Google Earth Engine (PlanetScope mosaic, Landsat). The data are customizable to different applications relevant to agriculture, land-use change, and natural resource management.

Methods for accuracy assessment differ according to whether the model output is a categorical variable (in the case of classification) or a continuous variable (in the case of regression). For example, a random forest algorithm can be used for classification or for regression, but the type of accuracy assessment and relevant accuracy metrics depend on which one is used. In both cases the assumption for assessing accuracy is that the validation dataset represents true values.

For classification analysis,[22] the analysis of accuracy is typically based on a confusion matrix or error matrix that compares the map classification with the reference data (Box 2). The confusion matrix evaluates the performance of the classification model by comparing the classes predicted from the classification algorithm to classes/category observed at the same locations in the reference data over the landscape. Different statistics are calculated based on the confusion matrix, including producer's accuracy, user's accuracy, and overall accuracy. These statistics should be reported along with the error matrix of estimated area proportion, the area of each class as determined from the map and estimated from the reference data, and standard error or confidence intervals that quantify the variability of the accuracy and area estimates.

[22] For additional guidance, we refer the reader to recent methodological studies that define good practices for accuracy assessment (Foody, 2002, 2009, 2021; Olofsson et al., 2013, 2014; Stehman and Foody, 2019).

Box 2. Illustration of accuracy assessment for a classified map

Table 1. Error matrix of the map classification and the reference data at the same sample location in number of pixels

		Reference data					
		Deforestation (A)	Forest gain (B)	Stable forest (C)	Stable non-forest (D)	Total sample size	User's accuracy
Map data	Deforestation (a)	140	0	30	30	200	0.70
	Forest gain (b)	0	90	30	30	150	0.60
	Stable forest (c)	20	0	2,880	300	3,200	0.90
	Stable non-forest (d)	40	20	250	6,140	6,450	0.95
	Total sample size	200	110	3,190	6,500	10,000	
	Producer's accuracy	0.70	0.82	0.90	0.94		0.93

The **producer's accuracy** reflects the error from a map maker's perspective by asking how often real features on the ground are correctly shown on the classified map. In other words, what is the probability that a certain land cover on the ground is classified correctly on the map? Omission errors (type I error) are calculated by reviewing the reference sites for incorrect classifications. Omission errors are the complement of the producer's accuracy.

Producer's accuracy = aA/ΣA = 140/200 = 0.70 or **Producer's accuracy = 1 - omission error**

Omission error = (bA+ cA + dA)/ ΣA = 0.30

The **user's accuracy** is from the map user's perspective and tells us how often the class on the map will actually be present on the ground. So, from the perspective of the user of the classified map, how accurate is the map? It is the complement of commission errors (type II error), which are calculated by reviewing the classified sites for incorrect classifications.

User's accuracy = aA/ Σa = 140/200 = 0.70 or **User's accuracy = 1 - commission error**

Commission error = (aB + aC + aD)/ Σa = 0.30

The **overall accuracy** is obtained by the number of points or pixels correctly identified between the map classification and the reference data. These are the diagonal elements identified in green in the table above. The overall accuracy is the sum of these pixels divided by the total number of pixels.

Overall accuracy = (140 + 90 + 2,880 + 6,140)/10,000 = 0.93

For regression analysis, a common performance measure used to quantify model accuracy is the root mean square error (RMSE) (Box 3), which is the square root of the average squared deviation between reference data and the predicted outcome values. If the reference data and/or model output contain many extreme values, one might consider using the mean absolute error (MAE), which is the average of the absolute difference between reference and predicted values (Pontius et al., 2008). The R-squared or adjusted R-squared is a (squared) measure of the linear correlation between the observed and predicted values (and squaring the value). In the context of regression, reference data are often used to estimate the confidence intervals based on the distribution of prediction errors. Practically, when validating regression predictions, it is useful to create a visual map of the residuals, in order to identify any spatial patterns or structure that could be potentially accounted for with additional controls.

Box 3. Accuracy assessment for regression

For continuous variables, typical regression-type models are evaluated on an independent dataset by comparing predicted and observed values at the same locations. Here, the root mean square error is used as a summary measure of the differences between values predicted by a model and the reference data values at the same locations.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(p_i - o_i)^2}{N}},$$

RMSE = root mean square error

p = predicted value
o = observed value
N = Number of sites
i = site i

The RMSE is one of the measures of performance frequently used in remote sensing. Note that the observed values are obtained from the reference data that has not been used for model training.

A rigorous accuracy assessment is a key part of any remote sensing analysis. It should be well documented to enhance the transparency and reproducibility of the method and results, and effectively incorporated into research outputs (e.g., publications and reports). Future work is still needed to establish clear guidance for accuracy assessment in all data contexts, in particular when using dense time-series analysis (Stehman and Foody, 2019). For impact evaluation, a rigorous well-documented accuracy assessment enhances the utility of research outputs for decision-makers and other end users.

IN BRIEF:

- *Accuracy assessment of remote sensing products is a key component of any mapping project.*
- *Validation reference data sets need to be independent of the training data sets and should be obtained from higher quality sources, e.g. higher resolution imagery or georeferenced field data.*
- *For classified maps, accuracy assessment should rest on a probabilistic sampling design, and the analysis of a confusion matrix to produce accuracy metrics, standard error/confidence intervals and biased-corrected area estimates.*
- *For regression analysis, the predictive performance of a continuous value map can be evaluated with different metrics, including the root mean square error. It is also recommended to produce a prediction error map and look for any spatial pattern.*
- *Accuracy assessment methodology and results should be published with the EO product.*

f. Implications of measurement error for impact evaluation

So far, we have addressed issues relevant to the acquisition and pre-processing of remote sensing data and the evaluation of model outputs produced from them. However, in the context of impact evaluation, this isn't the end of the story. Once these satellite-based outputs have been produced, they are typically incorporated into statistical analyses whose purpose is to estimate the causal effect of technology adoption on outcomes of interest. The next three sections address three key issues researchers should consider during this latter stage of causal inference analysis when using spatial data, beginning with the types of errors that can occur when evaluating impacts.

Most quantitative impact evaluations relevant for CGIAR use (linear) regression analysis, where the dependent variable y is an outcome of interest, and X is a set of explanatory variables that contains the measure of technology adoption. The regression coefficient associated with the technology adoption variable can be interpreted as representing the change in the outcome y associated with a unit change in the adoption variable. So, for example, if technology adoption is binary (0 = no adoption, 1 = adoption), then the coefficient captures the change in y that results from adopting the technology. However, for a host of reasons, when estimated using data from the real world, the predictive relationship identified by a regression is never perfect.[23] Even after careful methodological considerations to reduce error, some uncertainty from the measurement and natural variability will remain in the explanatory variables X, the outcome variable y, and thus in the relationship they form. Rigorous impact evaluation requires addressing uncertainty transparently.

Conceptually, uncertainty stems both from random errors, which are inversely proportional to precision, and from systematic errors (or bias), which refer to a lack of accuracy (Frey et al., 2006; Pelletier et al., 2013) (Figure 10). The two concepts are fully independent of each other.

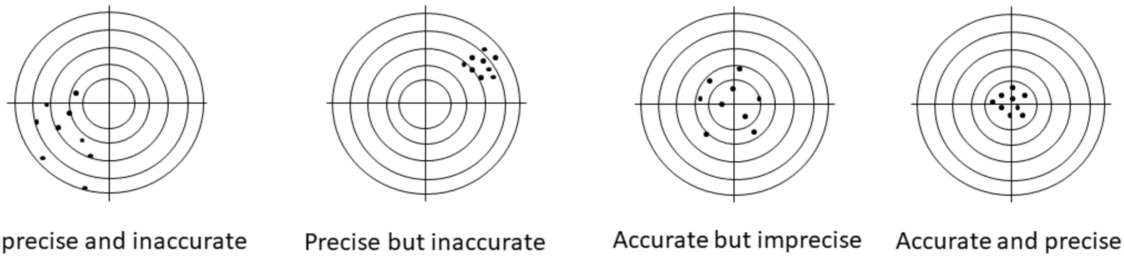


Figure 10. Illustration of precision and accuracy with the center of the bull eye being the true value. Precision describes the agreement among repeated measurements and characterizes the variation in measured observations. Accuracy is the agreement between the true value and the average of repeatedly measured estimates. Adapted from Frey et al. (2006).

Random errors, also known as "classical measurement error" or noise, are due to the variability in observations compared with the mean, which can be reduced by taking a sufficient number of observations. This source of error is easier to estimate through repeated measurements and describes both the precision of the measurement tool and the natural variability of the variable of interest. For example, random errors in the explanatory variables of a regression model may reduce the precision with which regression coefficients are estimated, which can affect the probability that a test of significance will pick up on an effect that is present. However, it will not introduce bias into the estimated coefficients. One can improve the statistical power by increasing the sample size, which reduces the effect of random errors.

Systematic errors, or lack of accuracy, may arise because of imperfections in conceptualization, models, measurement techniques, or other ways to make inferences from the data. Additional observations do not reduce systematic error, and so it can come to dominate the overall error. Like other data sources, EO data is affected by systematic error, and the accuracy assessment described in the earlier section serves to quantify it, under the assumption that the reference data used for validation represents the true value. For example, if we study the relationship between the adoption of improved seeds and a proxy for yield that is measured with a vegetation index that saturates after some threshold value, meaning that it stops registering variation after it reaches a certain value, the estimated relationship may be biased and may lead researchers to draw erroneous conclusions.

One prevalent type of systematic error is what is referred to in economics as non-classical measurement error (NCME). NCME occurs when there are systematic errors in the measurement of one or more of the variables included in a regression model. Correlated NCME happens when errors are correlated with the true value of the variable, with other variables included in the regression (such as the outcome), or with measurement errors in those variables (Abay et al., 2019). An example of correlated NCME comes from the Happy seeder case study in India. The team is interested in understanding the links between zero tillage (ZT), conventional tillage (CT), and crop residue burning. Residue burning produces smoke and haze that have the potential to affect the accuracy of the ZT-CT map by reducing

[23] Researchers often do not know (or observe) all the different factors that affect the outcome variable *y*. Some variables that produce variation in *y* are omitted from the set of explanatory variables *X*. As a result, the function of *X* estimated by the regression is able to predict *y* only imperfectly. The differences between the predicted values generated by the regression and the observed values of *y* are known as regression residuals. The residuals capture the variation in the outcome *y* that is not explained by the explanatory variables in *X*.

the quality of satellite data. If bias were introduced in the ZT-CT map because of smoke and haze, it could bias the estimated relationship between conservation agriculture and residue burning. For this reason, the research team is paying close attention to avoid any contamination by cloud or haze when mapping ZT-CT fields.

Recent studies have shown that ignoring correlated NCME can result in significantly biased regression coefficient estimates in studies with EO data (Alix-Garcia and Millimet, 2022). While emerging statistical techniques can potentially correct for this bias,[24] in general, practitioners of impact evaluations should mitigate NCME as much as possible. However, because geospatial impact evaluation often relies on modeled measures of outcomes or technology adoption, it may be difficult to avoid (correlated) NCME entirely. As a result, it is imperative for researchers to carefully assess the potential for correlated NCME in the data and modeled outputs they use. This problem is not an EO-specific problem, but it may be harder to track for practitioners that are less familiar these data sources. Practically, this involves scrutinizing the documentation and meta-data associated with any off-the-shelf EO-data products to understand how the data were produced and conducting a rigorous accuracy assessment of model outputs generated from EO-data using high-quality reference data.

Some projects have a complex workflow, involving the creation of various remote sensing–derived products or maps that are combined to generate the final output. This is especially the case for projects measuring both adoption and outcomes, but it can also be true in settings where spatially explicit data don’t exist a priori. When spatially explicit data are lacking, every input must be generated and mapped out in order to implement the impact evaluation. Since each map product comes with its own level of accuracy, errors can propagate or accumulate as the different maps are combined with one another during analysis. In this setting, it can be difficult to understand the nature of these emergent errors—for example, whether they are correlated with outcomes or technology-adoption measures—and therefore difficult to quantify how they contribute to the overall uncertainty of the results of an impact evaluation. One common approach is to use uncertainty propagation and Monte Carlo methods to generate overall uncertainty bounds. However, these calculations are dependent on adequate error estimation for each of the component pieces of the analysis. Accounting for higher-order model uncertainty of this nature may reduce the capacity for detecting the change in outcomes from technology adoption, but it supports a rigorous evaluation of impacts.

Another strategy to assess the nature of errors in a particular impact evaluation context is to zoom in to some areas where more accurate information can be obtained—for example, with higher-resolution images or/and field data. If there are concerns about systematic errors such as NCME, it is good practice to implement the regression analysis using data from a subset of the study area where more accurate reference data have been obtained. These results can then be compared with those obtained when the analysis is implemented using the full sample or with other subsets of the data where there is lower confidence in data quality. In assessing the overall consistency of the results, a researcher can identify potential sources of error in the data or methods used in the analysis and determine the level of confidence with which measured impacts should be interpreted (see Direct seed marketing program case study).

[24] For example, Alix-Garcia and Millimet (2022) develop extensions of estimators from Hausman et al. (1998) and Lewbel (2000), which offer improvements over current practices, even in the case of rare events data.

Case study: Direct seed marketing program in Ethiopia

Investigators: Johanne Pelletier, Solomon Alemu, Mira Korb, Travis Lybbert. Project status: In progress.

In 2011 the Ethiopian government launched the Direct seed marketing (DSM) program, which partially liberalizes seed markets by approving a range of certified seed providers for key crops. The program aims to improve smallholder farmers' access to improved seeds in order to increase yields. This study leverages the long time series of MODIS and Landsat data to measure changes in maize yields associated with the spatially heterogeneous rollout of DSM before and after its implementation.

Young maize farmer, Ethiopia.
Photo credit: C. Robinson/CIMMYT.

Thus far, the team has used phenological trend analysis of 250-m MODIS EVI data in locations observed in panel survey data (Mekonnen et al., 2021) to identify the peak of the maize growing season, used as a proxy for maize yield. Moving forward, the DSM team faces several challenges. There are no crop-type maps covering the period, so to identify maize crop areas the team will create annual maize maps using Landsat data. While the MODIS vegetation index at 250 m is ideal for phenological trend analysis, the resolution is too low to distinguish individual maize fields (the ~6 ha pixel captures mixed land uses). To rigorously assess all the potential sources of error that may have propagated in the workflow, they will zoom in by repeating the analysis using 30-m Landsat data and more detailed reference data. This will allow them to assess the overall consistency of their results and the level of confidence with which estimated impacts should be interpreted.

IN BRIEF:

- *Remote sensing data can contain both random (classical) and systematic (non-classical) errors, but the latter may bias the relationship being studied.*
- *Improving accuracy and adopting extended versions of estimators in impact models offer options to improve current practices.*
- *Quantifying overall uncertainty within a complex workflow through uncertainty propagation or Monte Carlo methods is recommended for rigorous analysis.*
- *Additional validation can be obtained by zooming into a subset of the study area where the consistency of results can be checked against more accurate information.*

g. Issues of scale and the modifiable areal unit problem (R)

In non-spatial impact evaluation, data analysis is typically done at a single unit of observation. The integration EO data provides options of analysis at different spatial scales. With increasing diversity and availability of EO data, including VHR images, we now have the choice to look at processes in much more detail and at higher frequency than it was possible before. The spatial (or temporal) resolution of satellite data imposes a filter through which a phenomenon is viewed. Since this spatial filter affects our understanding of relationships between variables of interest, it raises new questions about the appropriate scale to study a phenomenon.

Important concepts to consider when using geospatial data for impact evaluation are the terms "resolution" and "scale". The spatial resolution of an image refers to the size of a pixel in terms of ground dimensions. An image's resolution determines the smallest possible feature that can be detected on the image. The term "scale" can refer to different concepts. In a strict geographic sense, the scale is the ratio of a distance on a paper map to the distance on the ground.[25] It is commonly used to describe different phenomena or processes operating at different spatial and temporal scales, where scale can refer to extent and/or resolution (Levin, 1992). For example, a researcher might describe a program that distributes improved seeds to the heads of farm households as a household-scale intervention. Alternatively, one might describe the scale of a recent drought as taking place over the past several years.

For gridded EO data, the "measurement" scale of the data is the spatial resolution. But in the digital world, spatial products may represent data with a particular resolution that does not necessarily reflect the measurement resolution (Hijmans, 2021). For example, a 30-meter resolution soil carbon map does not indicate that soil samples covered an area of 30m by 30m, but rather suggests that the map was produced by a model that used 30-m resolution satellite imagery as an input. There is therefore soil carbon spatial variation that exist within the 30-m pixel (900 m^2) that is not captured, but the value consist in an estimate for the pixel area. Similarly, an EO product may be available at different spatial resolution. It is therefore important to understand the distinction between the resolution of the representation (data) and the resolution of the measurements or estimates, as the representation resolution can arise from aggregation or disaggregation (downscaling).

An old dilemma that pertains to the spatial scale of analysis is known in geography as the modifiable areal unit problem (MAUP). Fundamentally, the MAUP concerns the fact that statistical results are sensitive to changes in the spatial unit of analysis (Openshaw, 1984).[26] In other words, different levels of aggregation leads to different relationships between variables. The MAUP is important to consider when selecting the appropriate scale and unit of analysis for impact evaluation (Avelino et al., 2016) with satellite imagery since researchers can look at processes across different spatial scales. A key implication of the MAUP is that it is generally false to assume that outcomes or impacts observed at one scale will extrapolate accurately to a different one (Openshaw, 1984; Avelino et al., 2016).

The MAUP has two main dimensions: the scale effect and the zoning effect. The scale effect produces variation in statistics computed at different levels of aggregation. For instance, aggregating data to a higher level, a process called smoothing, leads to a loss of heterogeneity (Figure 11). The zoning effect refers to the change in the correlation between observations caused by the regrouping of data into different configurations at the same scale. Changes in the mean and variance of data that result from zoning effects are less predictable than those resulting from scale effects. That said, the overall effect of the MAUP generally has the largest impact on the variance of a spatial dataset. The magnitude of this effect depends on the level of spatial autocorrelation present in the data (Duque et al., 2018). We discuss this topic further in the next section.

[25] So, for example, if 1 cm on the map is equal to 100 m on the ground, the scale will be 1/10,000 (or 1:10,000).
[26] The MAUP dilemma applies equally to analyses using vector (e.g., polygons) and raster data (pixels).

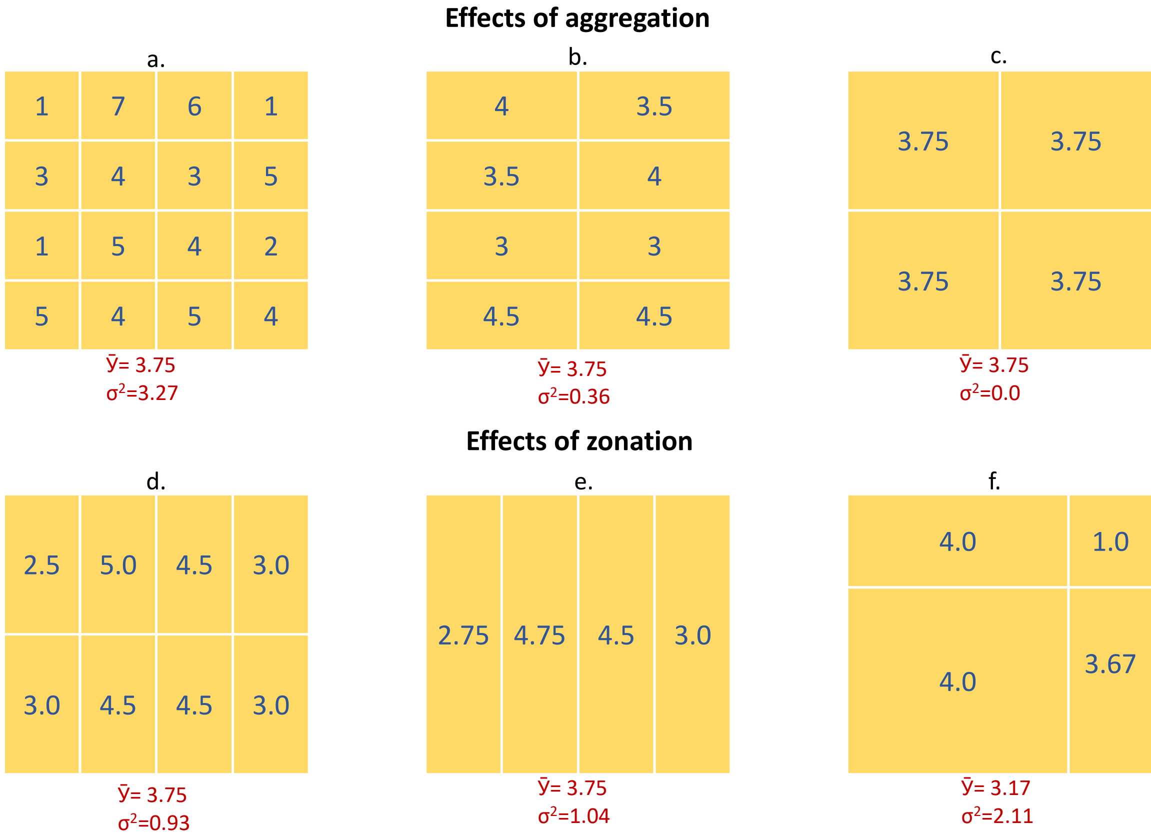


Figure 11. Panels a–c illustrate how a satellite image can be aggregated or resampled to a lower resolution. We see that the mean does not change, but the variance diminishes with aggregation (where $\bar{y}$ is the mean and σ^2 is the variance). Panels d–f show how different zonation or regrouping affects the variance. Source: Adapted from Dark and Bram (2007).

The issue of scale is not just about how we represent the data, it relates to the theoretical understanding of the data generating process. For socioeconomic variables, selecting the right unit of analysis requires that researchers have a good understanding of the social processes, including about the decision-making process and economic behavior that underlie the data. The general guidance for researchers on how to address this dilemma is to use a unit of analysis that matches the scale of the real-world phenomena that were measured to produce a dataset (e.g., physical, biological, or economic processes). Such an understanding is often gained through preliminary fieldwork and qualitative data collection.

When analyzing satellite imagery, the optimal resolution at which to conduct analysis depends on several factors, including the object(s) of interest in the image, the analytical methods being used, and the spatial structure of the image itself (Woodcock and Strahler, 1987). When the data-generating process is not clear, it is best practice to test relationships at different scales.[27] For example, a researcher analyzing a satellite image can plot how the local variance of a variable in each scene changes as a function of the changes in pixel dimensions by aggregating the data to lower resolutions and using the appropriate scale around the peak of variance.[28] It is also possible that data-generating processes operate at multiple different scales in the same system. So testing relationships at different scales helps

[27] For example, when analyzing vector data from a household survey, one can evaluate statistical relationships at the household level, as well as aggregating the data within administrative jurisdictions such as counties, states, or countries. For raster data, you can analyze outcomes at different resolutions (e.g., 5m, 10m, 30m, 100m, etc.)

[28] In this setting, one common decision rule is to use conduct the analysis at the resolution that maximizes the local variance of the variable of interest.

to examine how scale affects the statistical results.

More recently, Duque et al. (2018) have proposed a non-parametric statistical test, the S-maup, to measure the sensitivity of a spatial variable to scale effects. The S-maup test evaluates the changes in the distribution of a spatially disaggregated variable when it is aggregated into a given number of regions. Input parameters include the level of aggregation and the level of spatial autocorrelation in the variable. The S-maup test can be used to statistically compare two different aggregation levels. Crucially, it can also identify the maximum level of aggregation that preserves the distributional characteristics of the original variables and reduces the loss of information.

IN BRIEF:

- *The issue of scale is important when working with spatial data, both in theory because it serves as filter to observe a phenomenon and in practice because the statistical relationship between variables can change with a change in scale.*
- *The modifiable areal unit problem (MAUP) refers to the sensitivity of statistical results to changes in the spatial unit of analysis. It is risky to assume that relationship remains the same at different scale.*
- *We recommend to use a scale/resolution that represent the data generating process (e.g. decision-making process or behavior) as closely as possible.*
- *Other suggested methods are: running analysis at different scales, examining the local variance/resolution graphs and using the new S-maup statistical test in order to look at the sensitivity of the variable of interest to aggregation.*

h. Accounting for spatial correlation (R)

The key characteristic of spatial data is that they are structured in space. In general, observations that are closer in space are generally more similar, or more dissimilar, than randomly selected pairs of observations. Spatial correlation is a measure of similarity between nearby observations. When data are spatially correlated (and the nature of this correlation is known to a researcher), it is possible to predict the value of a variable at one location using information about its value in other nearby locations. This property is often exploited in order to interpolate the value of variables between locations where geolocated reference measurements were taken (e.g., using inverse distance weighting, or kriging).

Spatial correlation affects statistical testing, including for impact evaluation. Commonly used statistical methods assume that observations are fully stochastically independent from each other. However, when observations are spatially autocorrelated or spatially structured, they are not fully independent from each other. Spatial correlation implies that each new observation does not bring a full additional degree of freedom (also known as pseudo-replication). This is important because it creates biased impact estimates with unstable coefficients and unreliable significance tests. Fortunately, spatial correlation can be diagnosed, visualized, and accounted for in different ways.

Diagnosing and visualizing spatial correlation

Spatial structures in a response variable (or matrix of variables) can come from two distinct phenomena:

- induced spatial dependence, through external forcing by independent variables that are structured in space
- autocorrelation, through the internal structure of the response variable itself

In both cases, spatially correlated points near each other in space will show more similar values (positive correlation) or more dissimilar values (negative correlation) than will randomly located points.

Two main statistics can diagnose spatial correlation in univariate quantitative variables: the Moran's I (Moran, 1950), which works similarly to the Pearson correlation, and Geary's C (Geary, 1954), which is more similar to a distance measure (Borcard et al., 2011). For either method, a matrix of geographical distances among sites, also called a neighborhood matrix, is first constructed to calculate the spatial correlation coefficients.

The spatial correlation can be visualized relative to distance, or spatial lag, with a spatial correlogram—a plot that represents the spatial correlation values against the distance classes for one variable. A correlogram allows rapid characterization of the spatial correlation structure of a variable when it is combined with statistical tests (e.g., Moran's I or Geary's C). A typical pattern to observe on a spatial correlogram is a positive correlation at small distances that declines to negative values and becomes non-significant at a larger distance. In a correlogram, a test is performed for each lag (distance class). It is therefore important to apply correction of the p values for multiple testing (e.g., Bonferroni or Holm [1979] correction). Without correction, the overall risk of type I error is greatly increased.

In a multivariate context, spatial correlation can be assessed and tested for using a Mantel correlogram. For more information, we recommend that readers refer to multivariate analysis reference books by Legendre and Legendre (2012) and Borcard et al. (2011).

Adjusting for or predicting with spatial correlation

Spatial correlation creates problems for inference in regression models with spatial data. If there is spatial correlation in the residuals (when they should, in fact, be independent), it signifies that the model is mis-specified. This problem can be diagnosed and visualized as explained in the section above, creating a neighborhood matrix and using Moran’s I or Geary’s C to test for spatial correlation. Spatial correlation can sometimes be observed by mapping the residuals.

Once one has identified significant spatial correlation in regression residuals, different approaches can be used to adjust or address it in models. Spatial dependence can be integrated to regression models directly to understand the unique and joint variation explained by the spatial structure. For example, Moran eigenvectors maps (MEM) can be constructed and included in statistical models to understand the contribution of spatial structure to the explained variance. The spatial dependency can be used as weights in models using spatial lag model and spatial error model. Otherwise, the spatial structure can be broken up using randomization (see the Improved forages case study). In any case, impact evaluation using spatial data should use some statistical method to address issues of spatial correlation. Key references for spatial analysis include LeSage and Pace (2009) for spatial econometrics, Bivand et al. (2013) for spatial analysis in R, and Lovelace et al. (2019) for prediction, including spatial cross-validation.

IN BRIEF:

- *Spatial correlation is a measure of similarity (or dissimilarity) between nearby observations.*
- *Spatial correlation affects statistical testing in impact evaluation, because observations structured in space are not fully independent, thus creating a bias in impact estimates.*
- *Spatial correlation can be diagnosed with spatial correlogram and statistical testing.*
- *Different statistical approaches exist to address inference problems with spatial data, including by integrating or controlling for spatial dependence in regression models.*

Case study: Improved forages in the Ethiopian highlands

Investigators: Ignacio Sarmiento-Barbieri, Jason Sircely, Juan Camilo Cardenas Campo, Elias Zerfu, Rachid Laajaj. Project status: In progess.

Photo credit: ILRI/ Sonja Leitner

The Ethiopian highlands, which cover central and northern Ethiopia and form the largest contiguous area above 1,500m in Africa, are the setting for the improved forages project. Improved forages, whose intense rooting can reduce soil erosion, have been identified as win-win strategies with benefits for both livelihoods and soils. The case study aims to evaluate the impacts of the adoption of improved forages on environmental spillover using remotely sensed data on soil erosion, biomass production, and tree cover. The field team collected the perimeter of the main land uses at different sites where improved forages were adopted. The first analytical step for the project's team was to use land-use classification to map improved forage adoption, comparing different machine learning algorithms and satellite data sources. They used a k-shape time-series classifier on Sentinel-2 NDVI (10-m resolution) to capture the phenological stages of each vegetation type. One of the challenges with this analysis is that the improved forages area is small, covering only a limited number of pixels, compared with other land uses (grazing, cropland, and tree cover). Pixels under improved forages for one farm are assumed to be spatially correlated. To address spatial correlation and avoid undersampling problems for improved forages, the research team implemented a geographic block bootstrap unbalanced approach, balancing out the number of pixels within each land-use class. Block spatial cross-validation functions are used to iteratively build training data sets that can account for spatial correlation (Valavi et al., 2018). Spatial cross-validation can also reduce bias in the assessment of a model's predictive performance and helps avoid overfitting (Lovelace et al., 2019).

i. Transferability across landscapes

Many different factors contribute to the diversity of life on Earth. This diversity is rooted in ecological characteristics or agroecological zones (that, in turn, show different responses to human and natural disturbances), in agricultural landscape and cultural practices, and in other factors linked to social-ecological systems and their dynamics. This complex diversity means that innovations that work well in one context are not always equally successful in another context. This is also true for remote sensing methods to assess impacts. From our interviews with study teams, it became clear that the ability of EO tools to detect adoption or outcomes varied with the context. Leaving data constraints aside, there is deep context specificity to the systems of interest.

Certainly, several projects have noted variations in remote sensing product accuracy across changing landscape characteristics. For example, the remote sensing method developed in the Mekong River Delta in the AWD case study varies in its detection accuracy by soil type, despite Lovell's (2019) calibration of the images by carefully representing each of the three soil types in her sampling design.

Differences between soil types could come from differences in the interactions between the sensor and the soil type, but they could also be due to other characteristics that are associated with adoption, including farm size, farmers' income, access to farm laser leveling, and extension services. Another illustration comes from the STRV case study in Bangladesh, which obtained different levels of accuracy for rice mapping between different regions owing to differences in agricultural landscapes, ecological zones, and agricultural practices. The IBLI case study achieved better accuracy for land-cover classification in drier land than in more mesic areas, which usually include more diverse livelihood practices. When mapping zero-tillage/conventional tillage, different levels of accuracy obtained between Punjab/Haryana and Bihar could be explained not only by differences in the number of reference data points, but also by differences in field size, wealth, and crop types between the two regions (Conservation agriculture case study). If there are different levels of variance in different contexts, it can be helpful to understand which factors may limit accuracy.

Ultimately, this insight is relevant to the transferability of the remote sensing methods across different environments, including both spatial and temporal transferability. Some remote sensing methods have higher generalization capability than others. Methods that require a lot of operator intervention for training data selection and a posteriori labeling, for instance, can provide very good results for specific environments but are not always directly exportable to other geographies. For example, crop mapping with segmentation is time consuming because it requires a lot of operator intervention, which is not conducive to operational applications at large scale (Zhang et al., 2020). More robust detection methods typically have some automatic process that limits the need for operator intervention (e.g., rule-based methods, supervised temporal signal analysis [signal matching] for time-series analysis) and can rely on various satellite data sources from both passive and active sensors (Boschetti et al., 2017). Experience with common machine-learning algorithms has shown that it is better to avoid transferring models across regions, but rather to calibrate the algorithm with training data from the mapping region (Estes et al., 2018; Song et al., 2018; Elmes et al., 2020). One example of transferability is given by the Restoration of the commons case study in India.

Each outcome represents a unique Earth observation detection challenge requiring an appropriate methodology—sometimes simple, sometimes highly complex in terms of the required satellite data and processing steps. Different landscapes require some adaptation of the remote sensing methods and sometimes require more reference data for calibration. A clear note of caution for those interested in implementing geospatial impact evaluation: even when the method works in one place, there is no guarantee it will work elsewhere, at least not with the same level of accuracy. For this reason, piloting or testing approaches early on can help avoid disappointment when evaluating the impacts of interventions.

Case study: Restoration of the commons case study, India

Investigators: Karl Hughes, Tor-Gunnar Vagen, Leigh Winoweicki, Atul Dogra, Ruth Meinzen-Dick, Wei Zhang, Rahul Chatuvedi, Krister Andersson. Project status: In progress.

The Promise of the commons project looks at the environmental impacts of collective action efforts for land restoration in India. The project uses remote sensing to evaluate enhanced ecosystem services following the restoration of common lands by integrating both field measurements and remote sensing into an existing modeling framework, the Land Degradation Surveillance Framework (LDSF). This framework was originally developed in Mali and is now implemented with a consistent core methodology for monitoring land degradation (or restoration) in a network of 25 countries, with 60 sites (Figure 12), and for linking field measurements at these sites to remote sensing data. The project looks at carbon sequestration in vegetation and soils to quantify changes in ecosystem services associated with land restoration. Because of this existing modeling framework and consistent data collection design and methods, the project team was able to locally calibrate the LDSF with field data collection. The framework combines EO from different platforms, including MODIS, Landsat, RapidEye, and Sentinel-2 to assess change in soil organic carbon and vegetation cover in the context of this impact study.

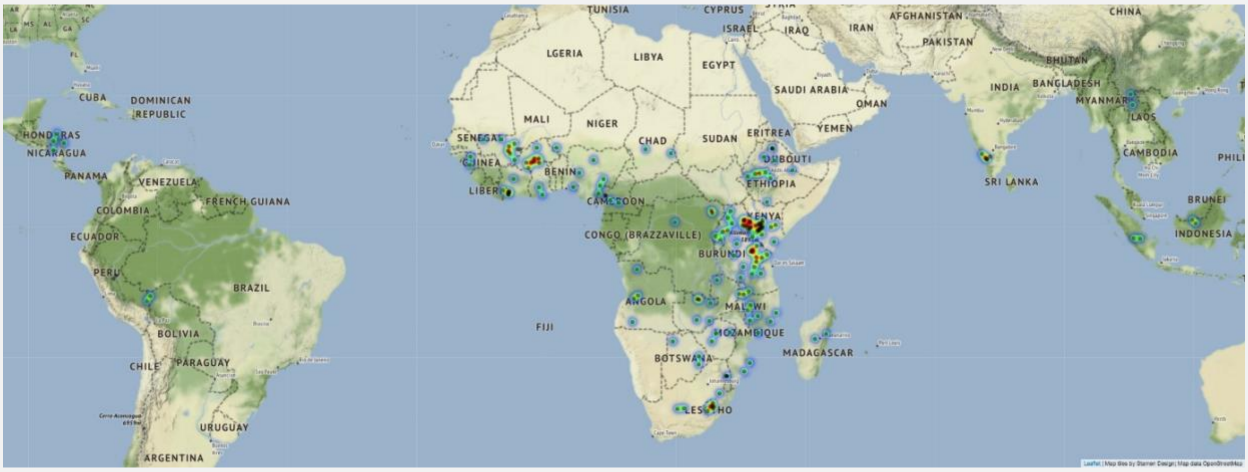


Figure 12. Network of 60 research sites in 25 countries (as of 2018), integrated in the Land Degradation Surveillance Framework. Source: http://landscapeportal.org/blog/2015/03/25/the-land-degradation-surveillance-framework-ldsf/

Clear acknowledgement by researchers of limitations and/or explanation of the factors that may affect the method's success in other areas can help to build transparency and confidence in the method. It is therefore essential to produce a realistic account of the limitations of the methods used and associated uncertainties.

Most case studies we reviewed here are limited to one innovation and one country. When a method works well in one area, it is valuable to test its transferability to another context. Remote sensing methods will be strengthened by cross-comparison between different settings. In this sense, applying the same method in different social-ecological contexts is a worthwhile endeavor. After the same method is effectively applied in different country contexts, there is a good chance that these remote sensing methods will become future standards for impact evaluation. Until then, remote sensing methods still need to be tested for impact evaluation and are likely to act as a complement to survey-based methods as this learning process continues.

IN BRIEF:

- ♦ *The level of accuracy of remote sensing products may differ between landscape contexts.*
- ♦ *Transparency about limitations of methods is important for expanding the field.*

- *Transferring methods and models to other areas requires testing using pilot study in order to adjust to the social-ecological context; some remote sensing methods have higher generalization capability too.*
- *The geospatial impact evaluation will be developed and strengthen by applying methods and algorithms to different contexts.*

7. EO for monitoring and impact evaluation within One CGIAR

The increasing availability of EO data underscores a new global impetus for monitoring and measuring progress toward attainment of the Sustainable Development Goals (SDGs). Under the 2030 Agenda for Sustainable Development, global leaders adopted 231 indicators for measuring, monitoring, and reporting progress toward targets identified in the SDGs. National statistical offices are encouraged to compile data on these indicators in order to assess their countries' performance and orient policy decisions. A recent analysis by the Group on Earth Observations (GEO) via the EO4SDG Initiative and the CEOS Ad Hoc Team on SDGs found that 34 of these outcome indicators are either directly or indirectly measurable using remotely sensed EO data (O'Connor et al., 2020).

How might remote sensing data related to these outcome indicators be integrated in future impact evaluation within One CGIAR? To start answering this question, we first show the relevance of the satellite-derived indicators assessed by O'Connor et al. (2020) to the five impact areas outlined in One CGIAR's 2030 Research and Innovation Strategy. We highlight some of the state-of-the-art applications of remote sensing data demonstrated in the literature. Second, we provide a summary assessment of the overall readiness and adequacy of specific EO data resources that practitioners may want to use or know about for measuring outcomes in impact evaluation. We showcase the results of this assessment in Annex 6, including weblinks to various EO data resources as well as readiness and adequacy scores for each indicator.

One CGIAR Impact Area #1: Poverty Reduction, Livelihoods, and Jobs

Black Marble data, NASA Earth Observatory

Poverty reduction, livelihoods, and jobs make up one of the five One CGIAR impact areas. EO data related to this impact area primarily concern SDG 1 (no poverty). Satellites have been used to construct poverty indicators in previous studies by combining Landsat 5, 7, 8, the Defense Meteorological Satellite Program (DSMP), and Visible Infrared Imaging Radiometer Suite (VIIRS) nightlights data, plus other covariates (Jean et al., 2016; Steele et al., 2017; Watmough et al., 2019; Yeh et al., 2020). Quickbird imagery has also been used to identify slum versus nonslum areas, which can be a good proxy for poverty levels (Montana et al., 2016). However, for these indicators both readiness and adequacy are considered only average. There is potential for EO to be used to track and target poverty, but no EO products exist to help distinguish poverty by age and gender. The technical capacity requirement is also high relative to other indicators. The temporal and spatial characteristics of these EO poverty indicators are yet not appropriate for all contexts, including for looking at the link between poverty and agriculture productivity indicators for instance. EO data and GIS can be applied to measure the distance to some basic services (e.g., road networks, waterways, main cities) (SDG indicator 1.4.1). For aspects of resilience to disaster, EO is best suited to evaluate damage to infrastructure. The economic impact on GDP, however, can only be inferred from visible damage (indicator 1.5.2).

One CGIAR Impact Area #2: Nutrition, Health, and Food Security

Nutrition, health, and food security make up another key impact area for One CGIAR. EO-based indicators related to this impact area concern SDG 2 (zero hunger), SDG 3 (good health and well-being), and SDG 6 (clean water and sanitation). Some indicators are also relevant to SDG 11 (sustainable cities and communities).

For food security indicators, a number of EO datasets can be used to quantify agricultural productivity through the estimation of crop area, crop yield, vegetation vigor, water stress, crop phenological development, and damage to crops (Atzberger, 2013; Huang and Han, 2014; Jin et al., 2017, 2018; Zhang et al., 2019; Jain et al., 2021; Kpienbaareh et al., 2021). Spatial resolution currently limits capacity to conduct farm-scale analyses of these indicators in smallholder systems. Yet there is also demonstrated experience that can be used to guide such efforts (Lobell et al., 2015, 2020; Jain et al., 2016, 2017; Azzari et al., 2017; Burke and Lobell, 2017; Jin et al., 2019). The proportion of agricultural area that is managed using improved and sustainable agricultural practices is also a relevant indicator for One CGIAR targets related to nutrition and food security. Here, household surveys remain the main source of data, but there is potential for EO to play a larger role in measuring the area under improved management practices. For example, EO data can be used to understand factors (e.g., environmental conditions) that may affect the adoption of farm management technologies, especially at the landscape level rather than at the individual farm level.

For health-related indicators, EO data are well suited for measuring air pollution, a significant global health issue. One active research interest in One CGIAR concerns air pollution produced from crop residue burning. EO datasets for measuring and monitoring air pollution are considered mature for a variety of different pollutants (e.g., PM2.5, CO2, CO, NOx, SO2) across multiple different scales (Anderson et al., 2012; van Donkelaar et al., 2015). Satellite instruments that gather data on airborne pollutants include Satellite Aura, Sentinel-5P, and spectroradiometers MODIS and MISR. EO can be used to detect the risk of dangerous level of air pollution against background atmospheric conditions. For example, fine particulate matter detection over cities is routine and operational. Recent studies have demonstrated that such data can be used to quantify the impact of agricultural fires on air pollution levels (Lelieveld et al., 2015) and, ultimately, human health outcomes (Ferguson and Govaerts, in review). One challenge to using EO data for measuring air pollution is that the technical capacity required for data processing and analysis is high, as illustrated by the Happy seeder case study in India. In addition, evaluating the impact of air pollution on health outcomes such as mortality typically requires additional non-satellite-based datasets—for example, administrative data from hospitals or health facilities, or individual health measures recorded in household surveys.

Case study: Happy seeder project, India

Investigators: Vijesh Krishna, Dhanyalekshmi Pillai, Meha Jain, ML Jat. Project status: In progress.

Photo: D. Vedachalam/CIMMYT

The Happy seeder is a direct-seeding technology that allows farmers to sow wheat under the residues from a previous crop. This technology reduces the need for farmers to burn crop residues prior to planting wheat, which is a common practice on irrigated rice-wheat systems in the Indo-Gangetic Plain. This project's goal is to measure how much the adoption of the Happy Seeder reduces air pollution associated with rice residue burning. The research team integrates various remote sensing products through the workflow. First, they analyzed Sentinel-2 images to map fields with conventional and zero tillage. Technology adoption was assessed through household survey data. EO-based measures of air pollution were produced through different steps. To identify fire locations involved a combination of MODIS Terra, Aqua, and Sentinel-2 imagery. Sentinel-5 (TROPOMI) and atmospheric transport modeling was then used to determine the change in greenhouse gas emissions (especially CO) associated with crop residue burning. In this ambitious study, the research team faced several challenges. First, it is hard to distinguish farmers using the Happy Seeder from farmers using other direct-seeder technologies. Second, the survey-based reference data may contain errors due to the misreporting of illegal crop residue burning practices. In addition, correctly attributing changes in greenhouse gas emissions and concentrations to crop residue burning requires complex modeling of local atmospheric conditions.

Crop residue burning, Punjab India, October 2021. Source: Google Earth.

EO data can also be used to construct various indicators related to water access and quality, which are significant determinants of human health. EO-based indicators of water quality include measures of turbidity and transparency, chlorophyll, temperature, suspended matter, and colored dissolved organic matter concentration. More sophisticated products can estimate particle size distributions and phytoplankton functional types, distinguish sources of suspended and colored dissolved matter, estimate water depth, and map types of heterogeneous substrates (Giardino et al., 2019). The Ocean and Land Color Instrument (OLCI) on Sentinel-3A and B, which is designed for water-quality monitoring, can provide near real-time detection of cyanobacteria in both inland and ocean settings. In general, these datasets exhibit variation in their level of maturity and confidence and should be carefully reviewed before use. For some indicators related to safe drinking water consumption and water pollution, EO can only partially support monitoring and modeling efforts. For example, EO-based methods for measuring water pathogens and salinity are known to have low accuracy. In addition, it is currently not possible to use EO data to assess water quality in groundwater or very small surface water bodies (e.g, less than 1 km^2 surface area).[29]

Another set of EO-based indicators may be used to evaluate dimensions of agricultural productivity and food security that pertain to water use efficiency and water stress. Satellites provide measures of many important parameters used in hydrological models that estimate water deficits (e.g., soil moisture, evapotranspiration, surface water extent, land-cover change, land use, surface temperature). There are also robust measurements for surface water extent and depth mapping of freshwater resources (e.g.,

[29] Sentinel-3 cannot map water bodies smaller than 1 km^2, while Sentinel-2 can monitor water quality of those as small as 150 by 150 meters, but with fewer parameters and high product uncertainty near land, so neither is adapted to evaluate water quality in small water bodies (Warren et al., 2021).

lakes, basins, reservoirs) based on optical sensors and scanning radiometry. The extent of irrigated agriculture, which typically accounts for the majority of freshwater withdrawals, can be mapped and used to estimate groundwater withdrawals estimated from land-cover products, vegetation productivity, and evapotranspiration parameters that can be constructed from satellite imagery.[30] Remotely sensed data contribute to measurements of water productivity (yield/m^3 of water consumed). Measures of agricultural water use efficiency are based on a variety of satellite- and non-satellite-based inputs (e.g., Bastiaanssen and Steduto's [2017] Water Productivity Score). Many of these indicators are also relevant for monitoring ecosystem health and dynamics, which is covered in the next subsection.

One CGIAR Impact Area #3: Environmental Health and Biodiversity

Environmental health and biodiversity constitute another key impact area for One CGIAR research. EO-based indicators related to this impact area concern primarily SDG 14 (life below water) and SDG 15 (life on land). Some are also related to SDG 6 (clean water and sanitation).

The monitoring and evaluation of technologies to protect and restore water-related ecosystems (SDG 14: life below water), specifically wetlands and coastal areas, can be well supported by remote sensing data (Pekel et al., 2016). Synthetic aperture radar (SAR), in particular L-band SAR, has been shown to penetrate below the vegetation canopies to detect the water surface of forested wetlands. This characteristic of SAR sensors allows for the effective detection of changes in the extent of wetlands, including in inundated riparian zones (Chapman et al., 2015; Muro et al., 2016). Beyond SAR, there are a number of other well-established EO-based methods for mapping the types and extent of vegetated wetlands (Rebelo et al., 2018). For example, Global Mangrove Watch monitors the extent of and changes in mangroves worldwide using a 25-m resolution global mosaic of JERS-1, ALOS, ALOS-2, and Landsat (Bunting et al., 2018).

In addition, environmental health impacts on water quality (SDG 6: clean water and sanitation) that arise from eutrophication due to agricultural runoff can also be assessed using EO-based indicators. For example, chlorophyll-a is a good proxy for eutrophication that can be easily monitored by satellites in most surface water bodies.[31] Satellite measures of colored dissolved organic matter and harmful algae or cyanobacteria can also be useful indicators of eutrophication.

Regarding terrestrial ecosystems, there are mature and reliable EO methods that can contribute to impact evaluation. For example, EO data have played a prominent role in the context of land-based climate change mitigation activities, including for monitoring forest dynamics since the advent of the UNFCCC REDD+ mechanism in 2013[32] (Hansen et al., 2013; GOFC-GOLD, 2016; GFOI, 2020) and, more recently, for natural climate solutions (Griscom et al., 2017, 2020; Bossio et al., 2020), which are mentioned in several One CGIAR initiatives. In this context, many EO methods and products have been developed to measure forest area change (Hansen et al., 2013; Song et al., 2018), forest degradation (Bullock et al., 2020), forest biomass (Saatchi *et al*., 2011; Baccini *et al*., 2012, 2017; Avitabile *et al*., 2016; Bouvet et al., 2018), and soil physical and chemical characteristics (Hengl et al., 2017, 2021; Sanderman et al., 2017). For impact evaluation, ancillary information is often still needed to specify technology adoption in forest management and conservation (e.g., georeferenced data on the boundaries of protected areas or the type and location of land management activities).

---

[30] Ferrant et al. (2017) find that agro-hydrological variables derived from Sentinel-1 and Sentinel-2 are crucial for quantifying the impacts of agriculture on water resources using spatially explicit agro-hydrological models of irrigation in India.

[31] It should be noted, however, that there is currently no consensus methodological approach for measuring chlorophyll-a in the coastal zone (O'Connor et al., 2020)

[32] REDD+ stands for Reduce Emissions from Deforestation and forest Degradation in developing countries, including forest carbon stocks enhancement, sustainable management of forests, and forest conservation.

One CGIAR Impact Area #4: Climate Change Adaptation and Mitigation

Climate adaptation and mitigation make up another key impact area for One CGIAR research. EO-based indicators relevant for this impact area are related primarily to SDG 13 (climate action) and SDG 7 (access to affordable and clean energy). Several studies have also demonstrated the potential of EO for measuring access to electricity using global and regional nighttime lights datasets (Elvidge et al., 2017; Román et al., 2018; Li et al., 2020) or to detect energy infrastructure at subnational scales. EO is commonly used to track the damage caused by disasters, including landslides (Casagli et al., 2016; Mondini et al., 2019; Solari et al., 2020), hurricanes (Vatsavai et al., 2011), fires (Giglio et al., 2018; Roy et al., 2019), floods (Twele et al., 2016; DeVries et al., 2020; Tellman et al., 2021), and droughts (West et al., 2019; Jiao et al., 2021; Chatterjee et al., 2022). These products and methods are incorporated into early warning networks and tools to support timely policy responses during crises, and they can be used in combination with population data to estimate the number of people affected by climate-related disasters. One key application is using EO data to measure greenhouse gas emissions, which is a rapidly advancing domain of remote sensing science. Technologies are currently available for mapping greenhouse gases at a relatively coarse scale, including water vapor, methane, carbon dioxide, nitrous oxide, ozone, ethane, propane, sulfur hexafluoride, chlorofluorocarbons, hydrofluorocarbons, and perfluorocarbons. New instruments (e.g., Sentinel-5P, TROPOMI, OCO-2 & 3), new applications of existing instruments (e.g., AVIRIS-NG), and other related sensors are generating new possibilities for measuring greenhouse gas emissions.[33] In general, using EO data in this way still requires high levels of technical competency.

Satellites also play a prominent role in collecting information on a group of variables known as Essential Variables (EVs), which are critical for terrestrial ecosystem monitoring.[34] EVs are biological, ecological, or physical parameters that experts have deemed critical for observing and monitoring changes in the Earth system. EVs have often been translated into global indicators because they are intended to be used in the measurement of progress toward global targets outlined in the SDGs. For example, Essential Climate Variables (ECVs) include measures of temperature, precipitation, and wind speed that are needed to parameterize climate models (Bojinski et al., 2014). More than half of the 55 identified ECVs are measured through satellite observations (O'Connor et al., 2020). Compared with ECVs, Essential Agricultural Variables (EAVs) have received less attention (Masó et al., 2020). In a recent review, Nakalembe et al. (2021) provide a list of EO-based EAVs that includes cropland masks (agricultural land), annual crop masks, crop-type masks, and EO-based crop-yield models and forecasts. Other satellite-based EAVs identified by the GEO Global Agricultural Initiative (GEOGLAM) include crop condition indicators and drought indicators.[35] The work of the GEO Biodiversity Observation Network (GEO BON) and the development of Essential Biodiversity Variables has progressed rapidly to create a large global network and community of practice involved in biodiversity observations. This network is also working to develop and promote the use of EOs for monitoring biodiversity (Skidmore et al., 2015, 2021) and, more recently, Essential Ecosystem Services Variables (EESV).

Apart from outcome indicators, the adoption and diffusion of many innovations resulting from research in One CGIAR could also possibly be observed and traced through EO with well-established methods, as for the Adoption of fertilizer tree case study. A full accounting of all the possibilities is beyond the scope of these guidelines, but opportunities exist for further integration of EO data and are predicted to expand with new development in remote sensing science and applications.

---

[33] A recent report from the Group on Earth Observation provides advice on existing and upcoming EO capabilities for greenhouse gas monitoring and open access data sources (GEO et al., 2021).

[34] We reviewed the most common sensors that could be used by researchers and practitioners within One CGIAR by focusing on those from the public domain in Annex 7. We identify their characteristics (spatial and temporal resolution, launch date, swath), their mission (the original goal for launching the instrument), and public/commercial status. The same imagery can be accessed through different data portals or APIs. By searching the name of satellite/sensors, the reader will find different download options.

[35] Crop condition indicators include NDVI, the Enhanced Vegetation Index (EVI), and the Vegetation Condition Index. Drought indicators include the Normalized Difference Water Index (NWVI), the Water Satisfaction Index (WSI), and the Standardized Precipitation Index (SPI), among others (Whitcraft et al., 2019).

Case study: Adoption of the fertilizer tree, Faidherbia albida, in Zambia

Investigator: Tor-Gunnar Vagen. Project status: Completed.

Photo: Charlie Pye-Smith/ICRAF

Photo: World Agroforestry Centre Archives

Remote sensing has been used to measure adoption of improved land management practices in the Eastern Province of Zambia. The project uses a standard geospatial method to map tree species distribution in order to evaluate adoption of the fertilizer tree, Faidherbia albida, among farmers in Eastern Zambia. Landsat 8 was used to map the presence or absence of the fertilizer tree on farmers' fields (Stevenson and Vlek, 2018). The presence of the fertilizer tree in farmers' fields is assumed to be a good indicator of adoption of this agroforestry practice because it reflects farmers' decision to keep these trees on their farms. A recent study has pushed for mapping the distribution of Faidherbia albida in Senegal based on publicly available Sentinel-2 time series data, at higher resolution (Lu et al., 2022), showing the progress in measuring adoption in different landscape contexts.

8. Conclusions

We are in a golden age for Earth observation applications, with free and open satellite remote sensing becoming available at higher spatial, temporal, and spectral resolution than ever before. There is also a whole landscape of organizations that are pushing to facilitate access and use of EO data for applications in monitoring and measuring progress toward the Sustainable Development Goals. As One CGIAR moves toward a more integrated systems approach, EO data can increasingly become a valuable asset for measuring the impacts of research and innovations. These guidelines provide a synthesis of the potential for using EO data for impact evaluation and spell out some common challenges that this exercise entails. We believe that by understanding both the state of the art and the limitations of remote sensing data and methods, researchers and practitioners will be better equipped to seize the opportunities that come with increased availability of and access to EO data.

Many developing countries lack historical data or national-scale monitoring programs. Satellite data can fulfill important data gaps in countries where people suffer the most from food insecurity, environmental degradation, and climate change impacts and can provide evidence to support the objectives of One CGIAR. This means that the potential of EO to contribute to development research is particularly relevant in many of the countries where One CGIAR is actively involved.

To leverage the full benefits of EO data within One CGIAR, it is important to understand the limitations to its expanded use. One of the most important constraint is the lack of reliable reference data for improving satellite-based model performance (Burke et al., 2021). Access to ground data is an important obstacle for fully leveraging remote sensing to measure impacts of development interventions in agriculture, natural resources, and the environment. The reality is that reference data are so crucial for training and validation in remote sensing analysis that the lack of accurate geocoded reference data was identified as the most important challenge by all of the case studies we reviewed.

One CGIAR has a unique capacity and competitive advantage to fill this ground data gap because of its in-country activities. In contrast to space agencies or other research institutions that use EO data, CGIAR Centers already have their ‘boots on the ground’ in developing countries, where reference data are lacking. For many projects and needs, this is a missed opportunity: people are already in the field, implementing other field survey methods, but not collecting the locational data required to leverage remote sensing data. The effort and added investment of acquiring more accurate GPS devices and of spending time to collect geocoded information are small in comparison with the potential benefit of integrating EO data. Collecting geocoded data opens the door for using remote sensing later, even many years after the project is completed. More elaborated measurements, including crop cuts, soil moisture, or soil carbon content, require greater investments and additional planning and logistics.

One CGIAR can play a strategic role in amplifying the contribution of EO data by helping to access ground reference data in developing countries for measuring impacts. Once satellite-based models are calibrated correctly, field data collection may become less expensive and time consuming. If remote sensing approaches are developed to reach a monitoring capacity, they can provide valuable feedback for adaptive management and facilitate rapid and cost-effective policy responses. Progress on how we measure impacts through the growing integration and advancement of EO methods and applications has the potential to not only benefit One CGIAR, but also support the evaluation of policies and interventions in countries where it is the most needed as well as of progress toward the SDGs.

9. Acknowledgements

This work was funded by SPIA, the Standing Panel on Impact Assessment of CGIAR. We are grateful to **Prof. Karen Macours** and **Dr. James Stevenson** for their unswerving support and guidance on this work.

We are indebted to the study teams of SPIA-funded studies for their time and availability to participate in interviews on which these guidelines are partly based. We warmly thank the following participants:

Dr. Tor-Gunnar Vågen and **Dr. Karl Hughes** (ICRAF), on the Faidherbia albida fertilizer tree, Zambia, and Restoration of the commons, India

Prof. Robin J. Lovell (New York University), on alternate wetting and drying (AWD), Vietnam

Dr. Anil Bhargava, on conservation agriculture (CA), India

Dr. Steve Wilcox and **Dr. Gerardo Soto** (Cornell University), **Dr. Nathaniel Jensen** (ILRI), and **Prof. Francesco Fava** (University of Milan), on index-based livestock insurance (IBLI), Ethiopia and Kenya

Dr. Dilini Abeygunawardane (Leibniz Institute of Agricultural Development in Transition Economies, IAMO), **Prof. Patrick Meyfroidt** (Université catholique de Louvain), and **Prof. Robert Heilmayr** (University of California, Santa Barbara), on the sorghum and millet upscaling project, Mali

Prof. Jenny Aker (Tufts University), **Prof. Kelsey Jack** and **Dr. Kendra Walker** (UC Santa Barbara), on the demi-lunes rainwater harvesting technique, Niger

Dr. Vijesh Krishna (CIMMYT), **Prof. Dhanyalekshmi Pillai** (Indian Institute of Science Education and Research), and **Prof. Meha Jain** (University of Michigan), on the Happy Seeder (direct-seeders for wheat sowing), India

Prof. Jeffrey D. Michler (University of Arizona) and **Dr. Renaud Mathieu** (IRRI), on stress-tolerant rice varieties (STRV), Bangladesh

Prof. Ignacio Sarmiento-Barbieri (Universidad de Los Andes), on improved forages, Ethiopia

We are grateful for the participation and minutes of a virtual workshop entitled "Remote Sensing for Impact Evaluation" that drew experts working on impact evaluation and remote sensing, organized by SPIA and emLab (Environmental Market Solutions Lab, at the University of California, Santa Barbara) in the fall of 2020.

10. References cited

Abay, K.A. *et al.* (2019) 'Correlated non-classical measurement errors,'Second best'policy inference, and the inverse size-productivity relationship in agriculture', *Journal of Development Economics*, 139, pp. 171–184.

Alix-Garcia, J. and Millimet, D.L. (2022) 'Remotely Incorrect? Accounting for Nonclassical Measurement in Satellite Data on Deforestation', *Journal of the Association of Environmental and Resource Economists* [Preprint].

Anderson H. Ross *et al.* (2012) 'Satellite-based Estimates of Ambient Air Pollution and Global Variations in Childhood Asthma Prevalence', *Environmental Health Perspectives*, 120(9), pp. 1333–1339. Available at: https://doi.org/10.1289/ehp.1104724.

Atzberger, C. (2013) 'Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs', *Remote Sensing*, 5(2), pp. 949–981.

Avelino, A.F.T., Baylis, K. and Honey-Rosés, J. (2016) 'Goldilocks and the Raster Grid: Selecting Scale when Evaluating Conservation Programs', *PLOS One*, 11(12), p. e0167945. Available at: https://doi.org/10.1371/journal.pone.0167945.

Avitabile, V. *et al.* (2016) 'An integrated pan-tropical biomass map using multiple reference datasets', *Global Change Biology*, 22(4), pp. 1406–1420.

Azzari, G. *et al.* (2021) 'Understanding the Requirements for Surveys to Support Satellite-Based Crop Type Mapping: Evidence from Sub-Saharan Africa', *Remote Sensing*, 13(23). Available at: https://doi.org/10.3390/rs13234749.

Azzari, G., Jain, M. and Lobell, D.B. (2017) 'Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries', *Remote Sensing of Environment*, 202, pp. 129–141.

Baccini, A. *et al.* (2007) 'Scaling field data to calibrate and validate moderate spatial resolution remote sensing models', *Photogrammetric Engineering and Remote Sensing*, 73(8), pp. 945–954.

Baccini, A. *et al.* (2012) 'Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps', *Nature Climate Change*, 2, pp. 182–185.

Baccini, A. *et al.* (2017) 'Tropical forests are a net carbon source based on aboveground measurements of gain and loss', *Science*, 358(6360), pp. 230–234.

BenYishay, A. *et al.* (2017) 'A primer on geospatial impact evaluation methods, tools, and applications', in *AidData Working Paper# 44*. AidData at William & Mary Williamsburg, VA.

Bivand, R.S., Pebesma, E.J. and Gómez-Rubio, V. (2013) *Applied Spatial Data Analysis with R*. 2nd edn. New York, NY: Springer. Available at: https://doi.org/10.1007/978-1-4614-7618-4.

Blankespoor, B. *et al.* (2021) 'Spatial anonymization: Guidance note prepared for the Inter-Secretariat working group on household surveys.' UN Inter-secretariat Working Group on Household Surveys Task Force on Spatial Anonymization in Public-Use Household Survey Datasets. Available at: https://unstats.un.org/iswghs/taskforces/documents/Spatial Anonymization Report submit01272021 ISWGHS.pdf.

Bojinski, S. *et al.* (2014) 'The concept of essential climate variables in support of climate research, applications, and policy', *Bulletin of the American Meteorological Society*, 95(9), pp. 1431–1443.

Borcard, D., Gillet, F. and Legendre, P. (2011) *Numerical Ecology with R*. Springer.

Boschetti, M. *et al.* (2017) 'PhenoRice: A method for automatic extraction of spatio-temporal information on rice crops using satellite data time series', *Remote Sensing of Environment*, 194, pp. 347–365. Available at: https://doi.org/10.1016/j.rse.2017.03.029.

Bossio, D.A. *et al.* (2020) 'The role of soil carbon in natural climate solutions', *Nature Sustainability*, 3(5), pp. 391–398.

Bouvet, A. *et al.* (2018) 'An above-ground biomass map of African savannahs and woodlands at 25m resolution derived from ALOS PALSAR', *Remote Sensing of Environment*, 206, pp. 156–173. Available at: https://doi.org/10.1016/j.rse.2017.12.030.

Brandt, M. *et al.* (2020) 'An unexpectedly large count of trees in the West African Sahara and Sahel', *Nature*, 587(7832), pp. 78–82.

Bullock, E.L. *et al.* (2020) 'Satellite-based estimates reveal widespread forest degradation in the Amazon', *Global Change Biology*, 26(5), pp. 2956–2969.

Bunting, P. *et al.* (2018) 'The global mangrove watch—a new 2010 global baseline of mangrove extent', *Remote Sensing*, 10(10), p. 1669.

Burivalova, Z. *et al.* (2015) 'Relevance of global forest change data set to local conservation: case study of forest degradation in Masoala National Park, Madagascar', *Biotropica*, 47(2), pp. 267–274.

Burke, M. *et al.* (2021) 'Using satellite imagery to understand and promote sustainable development', *Science*, 371(6535), p. eabe8628.

Burke, M. and Lobell, D.B. (2017) 'Satellite-based assessment of yield variation and its determinants in smallholder African systems', *Proceedings of the National Academy of Sciences*, 114(9), pp. 2189–2194.

Casagli, N. *et al.* (2016) 'Landslide mapping and monitoring by using radar and optical remote sensing: Examples from the EC-FP7 project SAFER', *Remote sensing applications: society and environment*, 4, pp. 92–108.

CGIAR System Organization (2021) *CGIAR 2030 Research and Innovation Strategy: Transforming food, land, and water systems in a climate crisis.* Montpellier, France: CGIAR System Organization. Available at: https://hdl.handle.net/10568/110918.

Chapman, B. *et al.* (2015) 'Mapping regional inundation with spaceborne L-band SAR', *Remote Sensing*, 7(5), pp. 5440–5470.

Chatterjee, S. *et al.* (2022) 'Soil moisture as an essential component for delineating and forecasting agricultural rather than meteorological drought', *Remote Sensing of Environment*, 269, p. 112833. Available at: https://doi.org/10.1016/j.rse.2021.112833.

Cochran, W.G. (2007) *Sampling techniques*. John Wiley & Sons.

Cunningham, D., Cunningham, P. and Fagan, M.E. (2019) 'Identifying Biases in Global Tree Cover Products: A Case Study in Costa Rica', *Forests*, 10(10), p. 853. Available at: https://doi.org/10.3390/f10100853.

Dark, S.J. and Bram, D. (2007) 'The modifiable areal unit problem (MAUP) in physical geography', *Progress in Physical Geography*, 31(5), pp. 471–479.

DeVries, B. *et al.* (2020) 'Rapid and robust monitoring of flood events using Sentinel-1 and Landsat data on the Google Earth Engine', *Remote Sensing of Environment*, 240, p. 111664. Available at: https://doi.org/10.1016/j.rse.2020.111664.

Duque, J.C., Laniado, H. and Polo, A. (2018) 'S-maup: Statistical test to measure the sensitivity to the modifiable areal unit problem', *PLOS One*, 13(11), p. e0207377. Available at: https://doi.org/10.1371/journal.pone.0207377.

Dwyer, J.L. *et al.* (2018) 'Analysis Ready Data: Enabling Analysis of the Landsat Archive', *Remote Sensing*, 10(9), p. 1363. Available at: https://doi.org/10.3390/rs10091363.

Elmes, A. *et al.* (2020) 'Accounting for Training Data Error in Machine Learning Applied to Earth Observations', *Remote Sensing*, 12(6), p. 1034. Available at: https://doi.org/10.3390/rs12061034.

Elvidge, C.D. *et al.* (2017) 'VIIRS night-time lights', *International Journal of Remote Sensing*, 38(21), pp. 5860–5879.

Estes, L. *et al.* (2018) 'The spatial and temporal domains of modern ecology', *Nature ecology & evolution*, 2(5), pp. 819–826.

Ferguson, J. and Govaerts, B. (In Review) 'Sustainable Agriculture, Residue Burning, and Urban Infant Mortality: Evidence from Mexico'.

Ferrant, S. *et al.* (2017) 'Detection of irrigated crops from Sentinel-1 and Sentinel-2 data to estimate seasonal groundwater use in South India', *Remote Sensing*, 9(11), p. 1119.

Filipponi, F. (2019) 'Sentinel-1 GRD Preprocessing Workflow', *Proceedings*, 18(1), p. 11. Available at: https://doi.org/10.3390/ECRS-3-06201.

Foody, G.M. (2002) 'Status of land cover classification accuracy assessment', *Remote Sensing of Environment*, 80, pp. 185–201.

Foody, G.M. (2009) 'The impact of imperfect ground reference data on the accuracy of land cover change estimation', *International Journal of Remote Sensing*, 30(12), pp. 3275–3281. Available at: https://doi.org/10.1080/01431160902755346.

Foody, G.M. (2021) 'Impacts of ignorance on the accuracy of image classification and thematic mapping', *Remote Sensing of Environment*, 259, p. 112367. Available at: https://doi.org/10.1016/j.rse.2021.112367.

Frey, H.C. *et al.* (2006) 'Uncertainties', in H.S. Eggleston et al. (eds) *General Guidance and Reporting, 2006 IPCC Guidelines for National Greenhouse Gas Inventories*. Hayama, Japan: Institute for Global Environmental Strategies (IGES).

Geary, R.C. (1954) 'The contiguity ratio and statistical mapping', *The incorporated statistician*, 5(3), pp. 115–146.

GEO (Group on Earth Observations), ClimateTRACE, and WGIC (World Geospatial Industry Council) (2021) 'GHG Monitoring from Space: A mapping of capabilities across public, private, and hybrid satellite missions'. GEO Secretariat, Geneva.

Gertler, P.J. *et al.* (2016) *Impact evaluation in practice*. Washington, D.C.: The World Bank. Available at: https://www.worldbank.org/en/programs/sief-trust-fund/publication/impact-evaluation-in-practice.

GFOI (2020) *Integrating remote-sensing and ground-based observations for estimation of emissions and removals of greenhouse gases in forests: Methods and Guidance from the Global Forest Observations Initiative*. Rome, Italy: Global Forest Observations Initiative Programme Office Food and Agriculture Organization of the United Nations.

Giardino, C. *et al.* (2019) 'Imaging spectrometry of inland and coastal waters: state of the art, achievements and perspectives', *Surveys in Geophysics*, 40(3), pp. 401–429.

Giglio, L. *et al.* (2018) 'The Collection 6 MODIS burned area mapping algorithm and product', *Remote Sensing of Environment*, 217, pp. 72–85.

GOFC-GOLD (Global Observation of Forest Cover and Land Dynamics) (2016) *A sourcebook of methods and procedures for monitoring and reporting anthropogenic greenhouse gas emissions and removals associated with deforestation, gains and losses of carbon stocks in forests remaining forests, and forestation.* Wageningen, The Netherlands: GOFC-GOLD Land Cover Project Office, Wageningen University.

Gorelick, N. *et al.* (2017) 'Google Earth Engine: Planetary-scale geospatial analysis for everyone', *Remote Sensing of Environment*, 202, pp. 18–27.

Grace, K. *et al.* (2019) 'Integrating environmental context into DHS analysis while protecting participant confidentiality: A new remote sensing method', *Population and development review*, 45(1), p. 197.

Griscom, B.W. *et al.* (2017) 'Natural climate solutions', *Proceedings of the National Academy of Sciences*, 114(44), pp. 11645–11650.

Griscom, B.W. *et al.* (2020) 'National mitigation potential from natural climate solutions in the tropics', *Philosophical Transactions of the Royal Society B*, 375(1794), p. 20190126.

Hansen, M.C. *et al.* (2013) 'High-Resolution Global Maps of 21st-Century Forest Cover Change', *Science*, 342, pp. 850–853.

Hausman, J.A., Abrevaya, J. and Scott-Morton, F.M. (1998) 'Misclassification of the dependent variable in a discrete-response setting', *Journal of econometrics*, 87(2), pp. 239–269.

Hengl, T. *et al.* (2017) 'SoilGrids250m: Global gridded soil information based on machine learning', *PLOS One*, 12(2), p. e0169748.

Hengl, T. *et al.* (2021) 'African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning', *Scientific Reports*, 11(1), p. 6130. Available at: https://doi.org/10.1038/s41598-021-85639-y.

Herold, M. (2009) *An assessment of national forest monitoring capabilities in tropical non-Annex I countries: Recommendations for capacity building*. Paper prepared for the Prince's Rainforests Project and the Government of Norway., p. 61.

Hijmans, R.J. (2021) *Spatial Data Analysis with R*. Available at: https://www.rspatial.org/raster/analysis/raster_analysis.pdf.

Hollstein, A. *et al.* (2016) 'Ready-to-Use Methods for the Detection of Clouds, Cirrus, Snow, Shadow, Water and Clear Sky Pixels in Sentinel-2 MSI Images', *Remote Sensing*, 8(8), p. 666. Available at: https://doi.org/10.3390/rs8080666.

Huang, J. and Han, D. (2014) 'Meta-analysis of influential factors on crop yield estimation by remote sensing', *International Journal of Remote Sensing*, 35(6), pp. 2267–2295.

Jain, M. *et al.* (2016) 'Mapping Smallholder Wheat Yields and Sowing Dates Using Micro-Satellite Data', *Remote Sensing*, 8(10), p. 860. Available at: https://doi.org/10.3390/rs8100860.

Jain, M. *et al.* (2017) 'Using satellite data to identify the causes of and potential solutions for yield gaps in India's Wheat Belt', *Environmental Research Letters*, 12(9), p. 094011. Available at: https://doi.org/10.1088/1748-9326/aa8228.

Jain, M. (2020) 'The benefits and pitfalls of using satellite data for causal inference', *Review of Environmental Economics and Policy*, 14(1), pp. 157–169.

Jain, M. *et al.* (2021) 'Groundwater depletion will reduce cropping intensity in India', *Science Advances*, 7(9), p. eabd2849.

Jean, N. *et al.* (2016) 'Combining satellite imagery and machine learning to predict poverty', *Science*, 353(6301), pp. 790–794.

Jiao, W., Wang, L. and McCabe, M.F. (2021) 'Multi-sensor remote sensing for drought characterization: current status, opportunities and a roadmap for the future', *Remote Sensing of Environment*, 256, p. 112313. Available at: https://doi.org/10.1016/j.rse.2021.112313.

Jin, X. *et al.* (2018) 'A review of data assimilation of remote sensing and crop models', *European Journal of Agronomy*, 92, pp. 141–152.

Jin, Z. *et al.* (2017) 'Mapping Smallholder Yield Heterogeneity at Multiple Scales in Eastern Africa', *Remote Sensing*, 9(9), p. 931. Available at: https://doi.org/10.3390/rs9090931.

Jin, Z. *et al.* (2019) 'Smallholder maize area and yield mapping at national scales with Google Earth Engine', *Remote Sensing of Environment*, 228, pp. 115–128.

Kotchenova, S.Y. *et al.* (2006) 'Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part I: Path radiance', *Applied Optics*, 45(26), pp. 6762–6774. Available at: https://doi.org/10.1364/AO.45.006762.

Kotchenova, S.Y. and Vermote, E.F. (2007) 'Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part II. Homogeneous Lambertian and anisotropic surfaces', *Applied Optics*, 46(20), pp. 4455–4464. Available at: https://doi.org/10.1364/AO.46.004455.

Kpienbaareh, D. *et al.* (2021) 'Crop Type and Land Cover Mapping in Northern Malawi Using the Integration of Sentinel-1, Sentinel-2, and PlanetScope Satellite Data', *Remote Sensing*, 13(4). Available at: https://doi.org/10.3390/rs13040700.

Legendre, P. and Legendre, L. (2012) *Numerical ecology*. 3rd edn. Amsterdam: Elsevier Science BV.

Lelieveld, J. *et al.* (2015) 'The contribution of outdoor air pollution sources to premature mortality on a global scale', *Nature*, 525(7569), pp. 367–371.

LeSage, J. and Pace, R.K. (2009) *Introduction to Spatial Econometrics*. 1st edn. New York: Chapman and Hall/CRC. Available at: https://doi.org/10.1201/9781420064254.

Levin, S.A. (1992) 'THE PROBLEM OF PATTERN AND SCALE IN ECOLOGY', *Ecology*, 73(6), pp. 1943–1967.

Lewbel, A. (2000) 'Identification of the binary choice model with misclassification', *Econometric Theory*, 16(4), pp. 603–609.

Li, X. *et al.* (2020) 'A harmonized global nighttime light dataset 1992–2018', *Scientific data*, 7(1), pp. 1–9.

Lillesand, T., Kiefer, R.W. and Chipman, J. (2015) *Remote sensing and image interpretation*. Hoboken, NJ: John Wiley & Sons.

Lobell, D.B. *et al.* (2015) 'A scalable satellite-based crop yield mapper', *Remote Sensing of Environment*, 164, pp. 324–333.

Lobell, D.B. *et al.* (2020) 'Eyes in the Sky, Boots on the Ground: Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis', *American Journal of Agricultural Economics*, 102(1), pp. 202–219. Available at: https://doi.org/10.1093/ajae/aaz051.

Lovelace, R., Nowosad, J. and Muenchow, J. (2019) *Geocomputation with R*. Boca Raton, FL: Chapman and Hall/CRC.

Lovell, R.J. (2019) 'Identifying Alternative Wetting and Drying (AWD) Adoption in the Vietnamese Mekong River Delta: A Change Detection Approach', *ISPRS International Journal of Geo-Information*, 8(7), p. 312.

Lu, T. *et al.* (2022) 'Mapping the Abundance of Multipurpose Agroforestry Faidherbia albida Trees in Senegal', *Remote Sensing*, 14(3), p. 662. Available at: https://doi.org/10.3390/rs14030662.

Masó, J. *et al.* (2020) 'Earth observations for sustainable development goals monitoring based on essential variables and driver-pressure-state-impact-response indicators', *International Journal of Digital Earth*, 13(2), pp. 217–235.

McRoberts, R.E. *et al.* (2016) 'Methods for evaluating the utilities of local and global maps for increasing the precision of estimates of subtropical forest area', *Canadian Journal of Forest Research*, 46(7), pp. 924–932.

McRoberts, R.E. *et al.* (2018) 'The effects of imperfect reference data on remote sensing-assisted estimators of land cover class proportions', *ISPRS Journal of Photogrammetry and Remote Sensing*, 142, pp. 292–300.

Mekonnen, D.K. *et al.* (2021) *The impact of Ethiopia's direct seed marketing approach on smallholders' access to seeds, productivity, and commercialization*. Washington, DC: International Food Policy Research Institute (IFPRI) (IFPRI Discussion Paper 1998).

Michler, J.D. *et al.* (2021) *Estimating the Impact of Weather on Agriculture*. Policy Research Working Paper 9867. Washingtion, D.C.: World Bank. Available at: http://hdl.handle.net/10986/36643.

Michler, J.D. *et al.* (2022) 'Privacy protection, measurement error, and the integration of remote sensing and socioeconomic survey data', *Journal of Development Economics*, 158, p. 102927.

Mondini, A.C. *et al.* (2019) 'Sentinel-1 SAR amplitude imagery for rapid landslide detection', *Remote Sensing*, 11(7), p. 760.

Montana, L. *et al.* (2016) 'Using satellite data to delineate slum and non-slum sample domains for an urban population survey in Uttar Pradesh, India', *Spatial Demography*, 4(1), pp. 1–16.

Moran, P.A. (1950) 'A test for the serial independence of residuals', *Biometrika*, 37(1/2), pp. 178–181.

Muro, J. *et al.* (2016) 'Short-term change detection in wetlands using Sentinel-1 time series', *Remote Sensing*, 8(10), p. 795.

Nakalembe, C. *et al.* (2021) 'A review of satellite-based global agricultural monitoring systems available for Africa', *Global Food Security*, 29, p. 100543. Available at: https://doi.org/10.1016/j.gfs.2021.100543.

O'Connor, B. *et al.* (2020) *Earth Observation for SDG: Compendium of Earth Observation contributions to the SDG Targets and Indicators*. Paris: European Space Agency.

Olofsson, P. *et al.* (2013) 'Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation', *Remote Sensing of Environment*, 129, pp. 122–131.

Olofsson, P. *et al.* (2014) 'Good practices for estimating area and assessing accuracy of land change', *Remote Sensing of Environment*, 148, pp. 42–57.

Openshaw, S. (1984) 'Ecological fallacies and the analysis of areal census data', *Environment and planning A*, 16(1), pp. 17–31.

Pekel, J.-F. *et al.* (2016) 'High-resolution mapping of global surface water and its long-term changes', *Nature*, 540(7633), pp. 418–422.

Pelletier, J., Gelinas, N. and Potvin, C. (2019) 'Indigenous perspective to inform rights-based conservation in a protected area of Panama', *Land Use Policy*, 83, pp. 297–307.

Pelletier, J., Martin, D. and Potvin, C. (2013) 'REDD+ CO2 flux estimation and reporting for early actions: dealing with uncertainty', *Environmental Research Letters*, 8, p. 034009.

Peter, B.G. and Messina, J.P. (2019) 'Errors in time-series remote sensing and an open access application for detecting and visualizing spatial data outliers using Google Earth Engine', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(4), pp. 1165–1174.

Pontius, R.G., Thontteh, O. and Chen, H. (2008) 'Components of information for multiple resolution comparison between maps that share a real variable', *Environmental and Ecological Statistics*, 15(2), pp. 111–142. Available at: https://doi.org/10.1007/s10651-007-0043-y.

Prudente, V.H.R. *et al.* (2020) 'Limitations of cloud cover for optical remote sensing of agricultural areas across South America', *Remote Sensing Applications: Society and Environment*, 20, p. 100414. Available at: https://doi.org/10.1016/j.rsase.2020.100414.

Rebelo, L.-M. *et al.* (2018) *The use of Earth Observation for wetland inventory, assessment and monitoring: An information source for the Ramsar Convention on Wetlands.* Ramsar Technical Report 10. Gland, Switzerland: Ramsar Convention Secretariat.

Román, M.O. *et al.* (2018) 'NASA's Black Marble nighttime lights product suite', *Remote Sensing of Environment*, 210, pp. 113–143.

Roy, D.P. *et al.* (2019) 'Landsat-8 and Sentinel-2 burned area mapping - A combined sensor multi-temporal change detection approach', *Remote Sensing of Environment*, 231, p. 111254. Available at: https://doi.org/10.1016/j.rse.2019.111254.

Saatchi, S.S. *et al.* (2011) 'Benchmark map of forest carbon stocks in tropical regions across three continents', *Proceedings of the National Academy of Sciences of the United States of America*, 108(24), pp. 9899–9904. Available at: https://doi.org/10.1073/pnas.1019576108.

Sabins Jr, F.F. and Ellis, J.M. (2020) *Remote Sensing: Principles, Interpretation, and Applications*. Long Grove, IL: Waveland Press.

Sanderman, J., Hengl, T. and Fiske, G.J. (2017) 'Soil carbon debt of 12,000 years of human land use', *Proceedings of the National Academy of Sciences*, 114(36), pp. 9575–9580.

Sannier, C., McRoberts, R.E. and Fichet, L.-V. (2016) 'Suitability of Global Forest Change data to report forest cover estimates at national level in Gabon', *Remote Sensing of Environment*, 173, pp. 326–338.

Skidmore, A.K. *et al.* (2015) 'Environmental science: Agree on biodiversity metrics to track from space', *Nature*, 523(7561), pp. 403–405. Available at: https://doi.org/10.1038/523403a.

Skidmore, A.K. *et al.* (2021) 'Priority list of biodiversity metrics to observe from space', *Nature Ecology and Evolution*, 5(7), pp. 896–906.

Solari, L. *et al.* (2020) 'Review of satellite interferometry for landslide detection in Italy', *Remote Sensing*, 12(8), p. 1351.

Song, X.-P. *et al.* (2018) 'Global land change from 1982 to 2016', *Nature*, 560(7720), pp. 639–643. Available at: https://doi.org/10.1038/s41586-018-0411-9.

Steele, J.E. *et al.* (2017) 'Mapping poverty using mobile phone and satellite data', *Journal of the Royal Society Interface*, 14(127), p. 20160690.

Stehman, S.V. and Foody, G.M. (2019) 'Key issues in rigorous accuracy assessment of land cover products', *Remote Sensing of Environment*, 231, p. 111199.

Stevenson, J., Macours, K. and Gollin, D. (2018) *The Rigor Revolution in Impact Assessment: Implications for CGIAR*. Rome, Italy: CGIAR Independent Science and Partnership Council (ISPC).

Stevenson, J. and Vlek, P. (2018) *Assessing the adoption and diffusion of natural resource management practices: Synthesis of a new set of empirical studies*. Rome, Italy: CGIAR Independent Science and Partnership Council (ISPC).

Tellman, B. *et al.* (2021) 'Satellite imaging reveals increased proportion of population exposed to floods', *Nature*, 596(7870), pp. 80–86. Available at: https://doi.org/10.1038/s41586-021-03695-w.

Tropek, R. *et al.* (2014) 'Comment on "High-resolution global maps of 21st-century forest cover change"', *Science*, 344(6187), pp. 981–981.

Twele, A. *et al.* (2016) 'Sentinel-1-based flood mapping: a fully automated processing chain', *International Journal of Remote Sensing*, 37(13), pp. 2990–3004.

Valavi, R. *et al.* (2018) 'blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models', *bioRxiv*, p. 357798.

van Donkelaar Aaron *et al.* (2015) 'Use of Satellite Observations for Long-Term Exposure Assessment of Global Concentrations of Fine Particulate Matter', *Environmental Health Perspectives*, 123(2), pp. 135–143. Available at: https://doi.org/10.1289/ehp.1408646.

Vatsavai, R. *et al.* (2011) 'Rapid damage assessment using high-resolution remote sensing imagery: Tools and techniques', in *Proceeding of the IEEE International Geoscience and Remote Sensing Symposium*. New York: IEEE, pp. 1445–1448.

Vermote, E. *et al.* (2016) 'Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product', *Landsat 8 Science Results*, 185, pp. 46–56. Available at: https://doi.org/10.1016/j.rse.2016.04.008.

Warren, M.A., Simis, S.G. and Selmes, N. (2021) 'Complementary water quality observations from high and medium resolution Sentinel sensors by aligning chlorophyll-a and turbidity algorithms', *Remote Sensing of Environment*, 265, p. 112651.

Watmough, G.R. *et al.* (2019) 'Socioecologically informed use of remote sensing data to predict rural household poverty', *Proceedings of the National Academy of Sciences*, 116(4), pp. 1213–1218.

West, H., Quinn, N. and Horswell, M. (2019) 'Remote sensing for drought monitoring and impact assessment: Progress, past challenges and future opportunities', *Remote Sensing of Environment*, 232, p. 111291. Available at: https://doi.org/10.1016/j.rse.2019.111291.

Whitcraft, A.K. *et al.* (2019) 'No pixel left behind: Toward integrating Earth Observations for agriculture into the United Nations Sustainable Development Goals framework', *Remote Sensing of Environment*, 235, p. 111470.

Woodcock, C.E. and Strahler, A.H. (1987) 'The factor of scale in remote sensing', *Remote Sensing of Environment*, 21(3), pp. 311–332.

Yeh, C. *et al.* (2020) 'Using publicly available satellite imagery and deep learning to understand economic well-being in Africa', *Nature Communications*, 11(1), p. 2583. Available at: https://doi.org/10.1038/s41467-020-16185-w.

Zhang, D. *et al.* (2020) 'A generalized approach based on convolutional neural networks for large area cropland mapping at very high resolution', *Remote Sensing of Environment*, 247, p. 111912. Available at: https://doi.org/10.1016/j.rse.2020.111912.

Zhang, J. *et al.* (2019) 'Monitoring plant diseases and pests through remote sensing technology: A review', *Computers and Electronics in Agriculture*, 165, p. 104943.

Zhang, Y. and Liang, S. (2020) 'Fusion of Multiple Gridded Biomass Datasets for Generating a Global Forest Aboveground Biomass Map', *Remote Sensing*, 12(16), p. 2559. Available at: https://doi.org/10.3390/rs12162559.

11. Acronyms

AGB	Aboveground biomass
AI	Artificial intelligence
ALI	Advanced Land Imager
ALOS	Advanced Land Observing Satellite
API	Application Programming Interface
ARD	Analysis-ready data
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
AVHRR	Advanced Very High Resolution Radiometer
AWD	Alternate wetting and drying
AWS	Amazon Web Services
BOA	Bottom-of-the-atmosphere
CA	Conservation agriculture
CDOM	Colored dissolved organic matter
CEOS	Committee on Earth Observation Satellites
CGIAR	Consultative Group on International Agricultural Research
CHIRPS	Climate Hazards Group InfraRed Precipitation with Station data
CIAT	International Center for Tropical Agriculture
CIFOR	Center for International Forestry Research
CSA	Climate-smart agriculture
CT	Conventional tillage
DAAC	Distributed Active Archive Centre
DEM	Digital Elevation Model
DHS	Demographic and Health Surveys
DLR	German Aerospace Centre
DSM	Direct Seed Marketing
EBV	Essential Biodiversity Variable
ECV	Essential Climate Variable
EDC	Euro Data Cube
emLab	Environmental Market Solutions Lab at UC Santa Barbara
EO	Earth observation
EO4SDG	Earth observation for Sustainable Development Goals
EROS	USGS Earth Resources Observation and Science Center
ESA	European Space Agency
ESDC	Earth System Data Cube
ET	Evapotranspiration
ETM+	Enhanced Thematic Mapper Plus
EU	European Union
EV	Essential Variable
EVI	Enhanced Vegetation Index
FAO	Food and Agriculture Organization of the United Nations
FAOSTAT	FAO Statistical Database
FAPAR	Fraction of Absorbed Photosynthetically Active Radiation

fCover	fraction of Green Vegetation Cover
FEWS	RFE Famine Early Warning Systems Rainfall Estimates
fPAR	fraction of Photosynthetically Active Radiation
FRA	Forest Resources Assessment
GCP	Global Precipitation Climatology Project
GCP	Ground control point
GDP	Gross domestic product
G-Econ	Geographically based economic data
GEDI	Global Ecosystem Dynamics Investigation Lidar
GEE	Google Earth Engine
GEO	Group on Earth Observations
GEO BON	Group on Earth Observations Biodiversity Observation Network
GEOBIA	Geographic Object-Based Image Analysis
GEOGLAM	Group on Earth Observations Global Agricultural Monitoring
GEO-GNOME	Global Network for Observation and Information in Mountain Environments
GEOSS	Global Earth Observation System of Systems
GFOI	Global Forest Observation Initiative
GFW	Global Forest Watch
GHG	Greenhouse gas
GHSL	Global Human Settlement Layer
GIE	Geospatial Impact Evaluation
GIS	Geographic Information System
GMIA	Global Map of Irrigation Areas
GMW	Global Mangrove Watch
GOCI	Geostationary Ocean Color Imager
GOSAT	Greenhouse Gas Observation Satellite
GPS	Global Positioning System
GPSDD	Global Partnership for Sustainable Development Data
GRACE	Gravity Recovery and Climate Experiment mission
GRUMP	Global Rural Urban Mapping Project
GSOC	Global Soil Organic Carbon
GSP	Global Soil Partnership
GUF	Global Urban Footprint
GWOS	Global Wetlands Observation System
HPC	High Performance Computing
HWSD	Harmonized World Soil Database
IA	Impact assessment
IAEG-SDGs	Inter-Agency Expert Group on SDG Indicators
IBLI	Index-based livestock insurance
ICAO	International Civil Aviation Organization
ICEP	Index of Coastal Eutrophication
ICRAF	World Agroforestry
IEA	International Energy Agency
ILRI	International Livestock Research Institute
InSAR	Interferometric Synthetic Aperture Radar

IOOS	Integrated Ocean Observing System
IPCC	International Panel on Climate Change
IRB	Institutional Review Board
IRRI	International Rice Research Institute
IRS	Indian Remote Sensing satellite
ISRIC	International Soil Reference and Information Centre
IUCN	International Union for Conservation of Nature
JAXA	Japanese Aerospace Exploration Agency
JERS	Japanese Earth Resources Satellite
JRC	Joint Research Centre
LAI	Leaf Area Index
LCCS	Land Cover Classification System
LDN	Land Degradation Neutrality
LiDAR	Light Detection and Ranging
LSMS	Living Standards Measurement Survey
LSWI	Land Surface Water Index
LULC	Land Use and Land Cover
MAUP	Modified Areal Unit Problem
MDG	Millennium Development Goal
MEA	Multilateral Environmental Agreement
MERIS	Medium Resolution Imaging Spectroradiometer
MISR	Multi-angle Imaging Spectro-Radiometer
MODIS	Moderate Resolution Imaging Spectrometer Sensor
MSI	Multi Spectral Imager
MSS	Multispectral Scanner
NASA	National Aeronautics and Space Administration
NCME	Non-Classical Measurement Error
NDVI	Normalized Difference Vegetation Index
NGO	Nongovernmental organization
NOAA	National Oceanic and Atmospheric Association
NPP	Net primary productivity
NSF	National Science Foundation (USA)
NSO	National statistic office
OBIA	Object-based Image
OCO	Orbiting Carbon Observatory
OLCI	Ocean and Land Color Imager
OLI	Operational Land Imager
PALSAR	Phased Array Synthetic Aperture Radar
PAR	Photosynthetically Active Radiation
PM	Particulate matter
PRODES	Programa de Cálculo do Desflorestamento da Amazônia (Deforestation in Brazil)
QA	Quality assurance
Radar	Radio detection and ranging
RCT	Randomized control trial

REDD+	United Nations Reducing Emissions from Deforestation and forest Degradation
RMS	Root mean square error
RMSD	Root-mean-square Deviation
RS	Remote sensing
SAO	Sentinel Alpine Observatory
SAR	Synthetic Aperture Radar
SDG	Sustainable Development Goal
SEPAL	System for Earth Observation Data Access, Processing and Analysis for Land Monitoring
SMAP	Soil Moisture Active Passive satellite mission
SMOS	Soil Moisture Ocean Salinity satellite mission
SNAP	Sentinel Application Platform
SOC	Soil organic carbon
SPIA	Standing Panel on Impact Assessment
SPOT	Système Pour l'Observation de la Terre (French satellite)
SRF	Strategy and Results Framework
SRTM	Shuttle Radar Topography Mission
STRV	Stress-tolerant rice variety
SWOS	Satellite-based Wetlands Observation Service
TIRS	Thermal Infrared Sensor
TOA	Top of atmosphere
TOMS	Total Ozone Mapping Spectrometer
TRMM	Tropical Rainfall Measuring Mission
TROPOMI	TROPOspheric Monitoring Instrument
TRWR	Total renewable freshwater resources
TWW	Total freshwater withdrawn
UAV	Unmanned aerial vehicle
UCS	Union of Concerned Scientists
UN	United Nations
UN CBD	Convention on Biological Diversity
UN ECOSOC	UN Economic and Social Council
UNCCD	UN Convention to Combat Desertification
UNFCCC	United Nations Framework Convention on Climate Change
UNFCCC	UN Framework Convention on Climate Change
UN-GGIM	UN Committee of Experts on Global Geospatial Information Management
UNISDR	United Nations International Strategy for Disaster Reduction
USA	United States of America
USGS	United States Geological Survey
UV	Ultraviolet
VCF	Vegetation Continuous Field
VCI	Vegetation Condition Index
VHR	Very High Resolution
VIIRS	Visible Infrared Imaging Radiometer Suite
VIS	Visible Spectrum
VNIR	Visible and Near-InfraRed

VPI	Vegetation Productivity Index
WaPOR	Water Productivity Open-access portal
WASH	Water, sanitation, and hygiene
WGGI	Working Group on Geospatial Information
WHO	World Health Organization
WOIS	Water Observation and Information
WRI	World Resources Institute
WSF	World Settlement Footprint
ZT	Zero tillage

12. Glossary

Term	Description
algorithm	In the context of remote sensing, algorithms generally specify how to determine higher-level data products from lower-level source data.
aperture	The diameter of an opening; the diameter of the primary lens or mirror of a telescope. A cross-sectional area of the antenna that is exposed to the satellite signal.
asynchronous	Not synchronized
absorption band	Wavelength interval within which electromagnetic radiation is absorbed by the atmosphere or other substances.
accuracy assessment	Compares a LULC, land use, or land cover classification map with a detailed, independently collected sample set named reference data or validation data. The validation data can be based on field observations and visual interpretation of higher spatial resolution imagery.
active sensor	Sensor that provides their own source of electromagnetic radiation to illuminate the terrain. Radar is one example.
atmospheric correction	Image processing procedures that compensate for effects of light scattering by the atmosphere in multispectral and hyperspectral data
band (channel)	A slice of wavelengths from the electromagnetic spectrum.
bit	A single digital unit of information.
background	Area on an image or the terrain that surrounds an area of interest or target.
backscatter	In radar, the portion of the microwave energy scattered by the terrain surface directly back toward the antenna.
bandwidth	The wavelength interval recorded by a detector. Also called spectral resolution.
beam	A focused pulse of energy.
bilinear	A common resampling technique that uses the value of the four nearest cells from the input image to interpolate the value of the georeference/orthrectified cell. Results in a smoother appearing image compared with the nearest neighbor technique.
binary	Numerical system using the base 2.
BRDF	Bidirectional reflectance distribution function. A measurement of the impact caused by different angles of incoming and reflected energy on the interpretation of features in the imagery.
C-band	Radar wavelength region from 3.8 to 7.5 cm.
calibration	The act or process of comparing certain specific measurements in an instrument with a standard.
change-detection images	A difference image prepared by digitally comparing images acquired at different times. The gray tones or colors of each pixel record the amount of difference between the corresponding pixels of the original images.
classification	Process of assigning individual pixels of an image to categories, generally on the basis of spectral reflectance characteristics.
color composite image	Color image prepared by combining three individual images in blue, green and red.
color IR image	NIR-red-green bands illuminated with RGB light.
contrast enhancement	Image processing procedure that improves the contrast ratio of images. The original narrow range of digital values is expanded to utilize the full range of available digital values.
contrast ratio	On a image, the ratio of reflectance between the brightest and the darkest

	parts of the image.
data set	A logically meaningful grouping or collection of similar or related data.
descending node	Direction a satellite is traveling relative to the Equator. A descending node implies a southbound Equatorial crossing.
digital	A means for encoding information in a communications signal by using bits (binary digits).
DEM	Digital elevation model. A general term that describes the Earth's topography with x and y coordinates (longitude and latitude) and z values specifying elevation at each x, y point.
detectability	Measure of the smallest object that can be discerned on an image.
difference image	Image prepared by subtracting the digital values of pixels in one image from those in a second image to produce a third set of pixels. This third is used to form the difference image.
digital image processing	Computer manipulation of digital image.
digital number (DN)	Value assigned to a pixel in a digital image.
digitization	Process of converting an analog display in a digital display.
distortion	On a image, changes in shape and position of objects with respect to their true shape and position.
electromagnetic	Relating to the interplay between electric and magnetic fields.
electromagnetic energy	Energy that travels at the speed of light in a harmonic wave pattern.
electromagnetic radiation	Energy transfer in the form of electromagnetic waves or particles that propagate through space at the speed of light.
electromagnetic spectrum	The entire range of radiant energies or wave frequencies from the longest to the shortest wavelengths--the categorization of solar radiation. Satellite sensors collect this energy, but what the detectors capture is only a small portion of the entire electromagnetic spectrum. The spectrum is usually divided into seven sections: radio, microwave, infrared, visible, ultraviolet, x-ray, and gamma-ray radiation.
enhancement	Process of altering the appearance of an image so that the interpreter can extract more information.
EO	Earth observation
error matrix	Used in accuracy assessment to compare the information from verification sites to information on the map for a number of sample areas. The matrix is a square array of numbers set out in rows and columns.
feature	An object or single entity that stores its geographic representation, both geometry and attribute data, which is typically a point, line, or polygon.
filter, digital	Mathematical procedure for modifying values of numerical data.
fluorescence	Emission of light from a substance stimulated by exposure the radiation from an external source.
full waveform	Lidar system that records the returned energy in a series of equal time intervals that yields a vertical summation of the returns from a pulse.
geographic coordinate system	A spatial reference system that uses latitude and longitude to locate features on the Earth surface.
Geographic Information system (GIS)	Integrated computer hardware, software, and data for capturing, storing, analyzing, and displaying geographically referenced information.

geometric correction	Image processing procedure that corrects spatial distortions in an image.
georeferencing	Remote sensing images and scanned maps are georeferenced to a standard map projection to enable their use with other geospatial layers in a GIS. Georeferencing adds x, y or latitude, longitude coordinates to each pixel in the image or scanned map. Also called rectification. Distortions due to topography are not corrected.
geostationary	Refers to satellites traveling at the angular velocity at which the Earth rotates; as a result, they remain above the same point on the Earth at all times.
GMT	Greenwich mean time. A universal 24-hoursystem for designating time.
GNSS	Global Navigation Satellite System. Collects accurate horizontal and vertical (x, y, z) coordinates for moving and fixed platforms. The US deployed the first GNSS named NAVSTAR Global Positioning System (GPS).
GPS	Global Positioning System deployed by the United States (first GNSS).
ground control point (GCP)	A geographic feature of known location that is recognizable on images and can be used to determine geometric corrections.
ground receiving station	Facility that records image data transmitted by a satellite, such as Landsat.
ground resolution	The ability to resolve terrain features on images.
ground swath	Width of the strip of terrain that is imaged by a scanner.
hyperspectral sensor	System that collects a continuous spectrum of reflectance at many narrow, contiguous, and closely spaced wavelength bands.
IFOV	Instantaneous field of view.
image	A portrayal of a scene or subject that is acquired by a digital system.
image swath	See ground swath.
imaging spectrometer	Synonym for hyperspectral scanner.
Incidence angle	In radar, the angle formed between a line normal to the target and another connecting the antenna and the target.
Instantaneous field of view	Solid angle through which a detector is sensitive to radiation. In a scanning system, the solid angle subtended by the detector when the scanning motion is stopped.
interferogram	In radar, images that record interference patterns created by superposing images acquired by two antennas that are separated by a short distance.
interpolation	The estimation of surface values at unsampled points based on known surface values of surrounding points.
interpretation	The process in which a person extracts information from an image.
land cover	Describes the materials (such as vegetation, rocks, or developed) that are present at the surface.
Landsat ETM+	Enhanced Thematic Mapper Plus. On Landsat 7.
Landsat MSS	Multispectral Scanner. On Landsat 1, 2, and 3.
Landsat OLI	Operational Land Imager. On Landsat 8.
Landsat TIRS	Thermal Infrared Sensor. On Landsat 8.
Landsat TM	Thematic Mapper. On Landsat 4 and 5.
land use	Describes how an area of land is used (such as crops, golf course, urban).
L-band	Radar wavelength region from 15 to 30 cm.
metadata	A set of descriptive information about the scene data contained in the archive. The information is sufficient for a user, during the process of scene query and selection, to determine at a minimum geographic coverage, date of collection sensor gain mode, time of acquisition, cloud cover, and other quality

	measurements.
mosaicking	The assembling of photographs or other images whose edges are cut and matched to form a continuous photographic representation of a portion of the Earth's surface.
multispectral	Sensing in usually 4 distinct wavelength bands (equivalent to colors, not all of which are visible to the human eye).
multispectral image	A remote sensing image created using data collected from more than one band.
map projection	A systematic representation of the curved surface of the Earth on a plane.
microwave	Region of the electromagnetic spectrum in the wavelength range from 0.1 to 30 cm.
mosaic	Composite image made by piecing together individual images covering adjacent areas.
minimum ground separation	Minimum distance on the ground between two targets at which they can be resolved on an image.
multipolarization image	Color image composited from radar images of three different polarizations.
multispectral system	Framing or scanning system that simultaneously acquires bands of different wavelengths that can have different bandwidths. The bands can be contiguous, separate, or overlap.
multispectral classification	Identification of land cover categories by digital processing of data acquired by multispectral scanners.
nadir	Point on Earth directly beneath a satellite, the opposite of zenith.
noise	Any unwanted and unmodulated energy that is always present to some extent within any signal.
nearest neighbor	A common and fast resampling technique appropriate for categorical or thematic raster data as it does not alter the values of the input cells.
near infrared (NIR)	Infrared region of the electromagnetic spectrum that includes wavelengths from 0.7 µm to 1 mm.
Normalized Difference Vegetation Index	A measure of vegetation vigor computed from multispectral and hyperspectral data.
orbit	The path of a body acted upon by the force of gravity.
orthoimage	Images that have been computer processed to remove geographic and topographic distortions
orthorectification	A geometric correction including a DEM in the rectification process to minimize input image distortions due to topographic relief.
panchromatic	Sensitive to all or most of the visible spectrum, between 0.4 and 0.7 micrometers.
passive Sensor	A type of remote sensing instrument, a passive sensor picks up radiation reflected or emitted by the Earth.
pixel	An abbreviation of picture element. The minimum size area on the ground detectable by a remote sensing device. The size varies depending on the type of sensor.
polar orbit	An orbit with an orbital inclination of near 90 degrees where the satellite ground track will cross both polar regions once during each orbit. The term describes the near-polar orbits of spacecraft.
polarimetric images	A sequence of radar images that record a range of polarizations from parallel-polarized to cross-polarized.

polarization	The direction in which the electrical field vector of electromagnetic radiation vibrates.
projected coordinate system	Systematic methods and mathematical transformations to transfer or "project" a map from the Earth's spherical surface onto a flat surface.
pulse	Short burst of electromagnetic radiation transmitted by a radar antenna.
quad polarization	Radar systems that transmit the signal as alternate pulses with H (horizontal) and V (vertical) polarizations and receive the signal simultaneously for both polarizations. HH, VV, HV, VH radar images are generated.
radar	Short for "radio detection and ranging", radar sends out short pulses of microwave energy and records the returned signal's strength and time of arrival.
radar altimetry	Nonimaging systems carried on aircraft and satellites to measure altitude with great precision. A pulse of microwave energy is transmitted vertically downward.
radiance	Digital number measured by the sensor that includes atmospheric scattering and target reflectance.
radiation	Propagation of energy in the form of electromagnetic waves.
radiometer	A device that detects and measures electromagnetic radiation.
radiometric	Relating to, using, or measured by a radiometer. The measurement of radiation.
radiometric resolution	The number of subdivisions, or bits, that an imaging system records for a given range of values. As the number of bits increases, the radiometric resolution increases.
raster data	An abstraction of the real world where spatial data is expressed as a matrix of cells or pixels, with spatial position implicit in the ordering of the pixels.
rasterize	The process of converting vector points, lines, and areas into raster image format.
raw data	Numerical values representing the direct observations output by a measuring instrument transmitted as a bit stream in the order they were obtained.
rectification	Another term for georeferencing.
red edge	The unique spectral line associated only with the spectra of vegetation that connects the absorption feature at the red wavelength with the reflectance feature at the NIR wavelength.
reflectance	Value if brightness in each pixel of a band that is corrected for the atmospheric scattering. Ratio of the radiant energy reflected by a body to the energy incident on it.
reflectance spectrometer	Instrument that records percent reflectance as a function of wavelength.
remote sensing	(1) In the broadest sense, the measurement or acquisition of information about some property of an object or phenomenon, by a recording device that is not in physical or intimate contact with the object or phenomenon under study. (2) Instruments that record characteristics of objects at a distance, sometimes forming an image by gathering, focusing, and recording reflected light from the Sun, or reflected radio waves emitted by the spacecraft.
resampling	Modifying the geometry of an image (which may be from either a remotely sensed or map data source). This process usually involves rectification and/or registration. More generally it refers to assigning values from one image to another image with empty cells that differs in geometry or/and resolution.
resolution	(1) A measure of the amount of detail that can be seen in an image; the size of the smallest object recognizable using the detector. (2) Intensity or rate of data sampling. In remotely sensed imagery, resolution is significant in four

	measurement dimensions: spectral, spatial, radiometric and temporal.
reflected IR	Electromagnetic region from 0.7 to 3.0 μm that consists primarily of reflected NIR and SWIR solar radiation.
reflectivity	Ability of a surface to reflect incident energy.
refraction	Bending of electromagnetic rays as they pass from one medium into a medium with a different index of refraction.
registration	Process of geometrically adjusting two images so that equivalent geographic points coincide. Also called spatial co-registration.
relief	Vertical irregularities of a surface.
repeat cycle	For Earth satellites in sun-synchronous orbits, the number of days between repeated observations.
roughness	In radar, the average vertical relief of small-scale irregularities of the terrain surface. Also called surface roughness.
satellite	Any body, natural or artificial, in orbit around a planet. The term is used most often to describe moons (natural satellites) and spacecraft (man-made satellites).
scene	Satellite image
spatial data	Any information about the location, shape of, and relationships among geographic features. This includes remotely sensed data as well as map data.
spatial resolution	The area on the ground that an imaging system (such as a satellite sensor) can distinguish.
spectral response	The relative amplitude of the response of a detector vs. the frequency of incident electromagnetic radiation.
spectrometer	An optical instrument that splits the light received from an object into its component wavelengths by means of a diffraction grating, then measuring the amplitudes of the individual wavelengths.
sun-synchronous orbit	An orbit in which a satellite is always in the same position with respect to the rotating Earth at the same time of day.
synchronous	The instantaneous alignment of two or more events in time. Events may occur at irregular intervals.
scale	Ratio of distance on a image to the equivalent distance on the ground.
scan line	Narrow strip on the ground that is swept by the IFOV of a detector in a scanning system.
scattering	Multiple reflections of electromagnetic waves by particles or surfaces.
scatterometer	Nonimaging radar device that quantitatively records backscatter of terrain as a function of incidence angle.
sensor	Device that detects electromagnetic radiation and converts it into a signal that can be recorded and displayed as either numerical data or an image.
SfM	Structure from motion. Multiray photogrammetry that uses multiple overlapping images of the same feature from different angles to generate a 3-D model.
shortwave IR (SWIR)	Reflected IR region from 0.9 to 3 μm that is employed in remote sensing.
signature	Set of characteristics by which a material or a object may be identified on a image or photograph.
spectral resolution	Range of wavelengths recorded by a detector. Also called bandwidth.

spectral reflectance	Reflectance of electromagnetic energy at specified wavelength intervals
spectrometer	A device designed to detect, measure, and analyze the intensity of electromagnetic radiation as a function of wavelength by using an optical grating or prism to disperse radiation.
supervised classification	Digital information extraction technique in which the operator provides training sites information that the computer uses to assign pixels to categories.
Synthetic aperture radar (SAR)	Radar system in which fine azimuth resolution is achieved by storing and processing data on the Doppler shift of multiple return pulses in such a way as to give the effect of a much longer antenna.
telemetry	(1) Radio signals from a spacecraft used to encode and transmit data to a ground station. (2) The science of measuring a quantity, transmitting the measured value to a distant station, and there, interpreting or recording the quantity measured.
thematic data	Thematic data layers in a data set are layers of information that deal with a particular theme. These layers are typically related information that logically goes together.
thermal infrared	Electromagnetic radiation with wavelengths between 3 and 25 micrometers.
target	Object on the terrain of specific interest in a remote sensing investigation.
temporal resolution	the time interval between successive images.
terrain	Surface of the Earth.
texture	Frequency of change and arrangement of tones on an image.
TIR image	Image acquired by a scanner that records radiation within the TIR region.
training site	Area of terrain with known properties or characteristics that is used in supervised classification.
unsupervised classification	Digital information extraction technique in which the computer clusters pixels into natural groupings based on the spectral characteristics of the pixels with no instructions from the operator except for setting basic parameters.
vector data	Vector data, when used in the context of spatial or map information, refers to a format where all map data are stored as points, lines, and areas rather than as an image or continuous tone picture. These vector data have location and attribute information associated with them.
visible radiation	The electromagnetic radiation that humans can see as colors. The visible spectrum is composed of wavelengths between 0.4 to 0.7 micrometers. Red is the longest and violet is the shortest.
wavelength	The distance from crest to crest, or trough to trough, of an electromagnetic or other wave. Wavelengths are related to frequency: The longer the wavelength, the lower the frequency.
Worldwide Reference System (WRS)	A global indexing scheme designed for the Landsat Program based on nominal scene centers defined by path and row coordinates.
X-band	A nominal frequency ranges from 12.5 to 8 GHz (2.4 to 3.75 cm wavelength) within the microwave (radar) portion of the electromagnetic spectrum. X-band is a suitable frequency for several high-resolution radar applications and is used for both experimental and operational airborne systems
zenith	The point on the celestial sphere directly above the observer. Opposite the nadir.

Annex 1. Impact evaluation in a nutshell

Causal impact evaluation is a specific type of evaluation that typically seeks to establish and quantify how an intervention or other exogenous change affects an outcome.[36] At its core, impact evaluation is a **causal inference** problem—it seeks to determine the causal relationship between a change in a variable X (which can be an intervention, an innovation, or a policy change) and the change in the outcomes of interest Y.

Impact evaluations can be structured around the following type of research question: What is the impact or causal effect of this innovation on an outcome of interest? For instance, a project's interest can be whether the adoption of zero tillage increases maize yields among smallholder farmers. Adoption is a precondition for generating impacts at scale and provides a signal that the innovation generates sufficient interest from and/or benefits to the adopters, but measuring adoption does not constitute impact evaluation per se. In other contexts, the interest might be comparing different ways of implementing a program or different technologies on the same outcomes, with this question in mind: Which one is most effective?

A key concept of impact evaluation is **attribution**. Impact evaluation is concerned with the changes in outcomes that can be directly attributable to or caused by a particular intervention, innovation, or policy change.

To answer questions about causality and attribution, impact evaluation brings empirical research tools from economics and other social sciences. These methods were in turn inspired by experimental methods developed in the medical field to test the effectiveness of new treatments. Impact evaluation methods therefore adopt a similar vocabulary.

A key concept when studying causality is the **counterfactual**. The counterfactual refers to the world as it would have been in the absence of the intervention studied. The ideal way to identify impact would be to compare this counterfactual world to the actual world. Since it is impossible to simultaneously observe the world where the intervention took place and the counterfactual world where it did not, impact evaluation researchers have to find a way to simulate a valid counterfactual. Just as in medical experiments, this is often done by using a control group. A crucial step is thus to define a control group (not receiving treatment) that is in all respect similar to the treatment group.

Expressed differently, the key question of an impact evaluation can be described by this simple formula:

$$\alpha = (\mathrm{Y} \mid \mathrm{I} = 1) - (\mathrm{Y} \mid \mathrm{I} = 0),$$

where α is the causal impact of an intervention I on an outcome Y. The causal impact is calculated as the difference between the outcome Y with the intervention taking place $\mathrm{I} = 1$ and the same outcome in the absence of the intervention $\mathrm{I} = 0$. Ideally, one would compare the outcome on the same unit of observation, with and without the intervention taking place. In other words, a counterfactual is an estimate of what the outcome (Y) would have been for a program participant in the absence of the intervention (I). However, because we have no means to obtain information for two parallel universes, one with and one without the intervention taking place for the exact same unit, period, and conditions, constructing a robust counterfactual requires great care.

To estimate a proper counterfactual, the impact evaluator will generally estimate average impacts of an intervention on a population by developing a sampling design and using statistical tools to define the treatment group and the control group. A valid counterfactual will possess the same characteristics as the treatment group, apart from the fact that the units in the control group are not involved in the intervention. Three main characteristics can qualify a valid control group. First, the control group needs to have very similar average statistical characteristics to the treatment group before the start of the intervention. Second, both groups should be expected to react the same way. Third, the treatment and

---

[36] This section was inspired largely by *Impact Evaluation in Practice* (Gertler et al., 2016). We recommend this reading for those who would like a more complete, nontechnical overview of impact evaluation.

control groups must not be exposed to other interventions during the implementation period under evaluation.

The strongest way to create a valid counterfactual is by using randomization to assign the treatment. This selection process is generally done with a two-step sampling design. First, a sample is selected randomly from the population of eligible units, to be used as the evaluation sample. If this evaluation sample is large enough, it can be representative of the general population of all eligible units and preserve its characteristics (similar means and standard deviation). This ensures that the results can be generalized to the whole population of eligible units and thus have **external validity**.

Second, part of this evaluation sample is randomly assigned to the treatment group and another part to the control group. To ensure that the evaluation has **internal validity**, both groups should share identical statistical characteristics before the project's implementation so that the control group constitutes a valid counterfactual. After the intervention, it is the difference in outcomes between the treatment group and control group that can be attributed to the impact of the intervention, free of confounding factors.

Measuring impacts by making a before-and-after comparison of the intervention for the same treatment group would not produce a valid measure of impacts because there would be no true counterfactual. Similarly, comparing enrolled and non-enrolled groups—that is, assignment to a group is not random but based on groups' self-selection related to the outcome of interest—is also not valid because the group that enrolls is not comparable to the group that does not enroll. This last case can result in **selection bias** if enrollment is correlated with outcomes, which can occur when the control group was not part of the eligible population or decided not to participate.

Depending on data availability, different statistical methods can be employed to evaluate impacts. These include randomized assignment of treatment and control groups, as mentioned, but also quasi-experimental methods, such as regression discontinuity design, instrumental variables, or differences-in-differences. Quasi-experimental methods try to evaluate the causal impact of an intervention like experiments but without random group assignment (nonrandom criteria). Key to the credibility of a quasi-experimental design is the argument that the study design creates a division between treatment and control groups that make them, by all relevant means, comparable in the specific setting of the study. Detailed information on these methods is beyond the scope of this document.

Impact evaluations can be performed in a **prospective** manner (developed at the same time as the program design and implementation) or in a **retrospective** manner (developed after the program has been implemented). Prospective impact evaluations are generally more desirable because they are more likely to produce valid counterfactuals, leading to more robust evaluation results. The prospective approach can create better counterfactuals for three main reasons. First, this approach can provide baseline data before the project is implemented to measure preintervention outcomes. Second, there is flexibility to define the measures of the intervention's impact and make sure those measures are collected. Third, the assignment to treatment and control groups can be done before the program is implemented, ensuring random assignment and similarity between groups, apart from the change caused by intervention. Among the case studies that we reviewed in this document, the Demi-Lunes case study is a prospective geospatial impact evaluation study.

In contrast, retrospective evaluations aim to assess the impacts of programs after their implementation. Remote sensing data are well suited for retrospective evaluation because images are routinely collected independently of the intervention. However, with this retrospective strategy, it is necessary to rely on existing data to assess programs, and the treatment and control groups are generated ex post. The flexibility to define a valid counterfactual is more limited, and the measures of the program's success are dependent on available information. Thus, the feasibility of retrospective evaluation depends on the availability of data with sufficient spatial and temporal coverage of the treatment and control groups, before and after implementation. Retrospective evaluations often use quasi-experimental methods (BenYishay et al., 2017). Retrospective evaluations also generally build on stronger assumptions, and the credibility of the impact estimate relies on how well the study makes the argument that these assumptions hold and how well it is executed. Selecting good performance indicators that well represent the outcomes of interest is also essential to provide robust and credible evidence of impact.

Annex 2. Remote sensing in a nutshell

Remote sensing is a process of acquiring information about the physical characteristics of an area, object, or phenomenon from a distance. It is also described as "the science of acquiring, processing and interpreting" the interaction between electromagnetic energy and matter, such as light, heat, and microwaves (Sabins and Ellis, 2020). Other definitions are broader and include all acquisition of information from a distance, including acoustics measurements or sonar in the water. Here, we limit our discussion to vision-based remote sensing.

Acquiring remote sensing data involves an instrument or sensor that is comparable to a camera mounted on a platform, typically a satellite or aircraft. The sensor measures the electromagnetic radiation that is reflected or emitted by the target. The features of the sensor or platform define the spatial resolution, revisit frequencies, signal-to-noise ratio, and spectral capabilities of the images collected.

Satellites have three main types of orbits to observe the Earth. First, a polar-orbiting satellite has an orbit inclined perpendicular to the equator plane. This inclination provides a view of the entire globe, including regions that are hard to reach (e.g., the poles). The polar-orbiting satellites can be ascending (moving south to north) or descending (moving north to south) when they pass the equator. They are considered sun-synchronous when for each cycle they pass over the same location at the same solar time. Second, a non-polar-orbiting satellite orbits the Earth at low altitudes and generally provides a limited range of latitudes. Third, a geostationary satellite follows the rotation of the Earth to collect information continuously in one spot. Examples include weather satellites.

Sensors (instruments carried by satellites or aircraft) are classified into passive and active sensors. **Passive sensors**, which are used for most remote sensing applications, rely on the sun's energy as the main source of illumination and measure the energy that is reflected back to them from the Earth. The energy reflected back includes the visible, infrared, thermal infrared, and parts of microwave regions of the electromagnetic spectrum, and is used to measure, among other things, vegetation, land surface temperature, clouds and aerosols. These wavelengths cannot penetrate cloud cover, so passive sensors have limitations for use in tropical areas where dense cloud cover is frequent. **Active sensors** provide their own source of energy. The radio detection and ranging (radar) instruments cover the microwave portion of the electromagnetic spectrum and are commonly denoted by letters, including the X-band, C-band, L-band, and P-band. They direct energy at the target and record the returned energy (and time) and are insensitive to clouds. Light detection and ranging (lidar) uses laser light beams with higher-frequency (shorter wavelength) pulses than radar and is therefore sensitive to cloud cover. These active systems are used to measure, among other things, forest structure, ice, precipitation, wind, and the vertical profile of aerosols.

Remote sensing expands the visible range of the human eye by converting the range of electromagnetic spectrum that is invisible to our eyes (thermal, radio, micro wave) into images that we can observe and analyze to detect and monitor changes at the Earth's surface. **Electromagnetic energy**, which is produced by the vibration of charged particles, travels at the speed of light through space and through the atmosphere in the form of waves. Radiation of higher electromagnetic energy has shorter wavelengths and higher frequency (e.g., ultraviolet rays with wavelengths between 0.03 to 0.4 μm) and will often be intercepted by gases in the atmosphere (e.g., water vapor, carbon dioxide, ozone layer). Radiation with low electromagnetic energy has larger wavelengths and low frequency (the microwave region 0.1 to 100 cm) and can pass through clouds (e.g., the microwave radar). Visible light sits in the middle of that range of long to shortwave radiation in a very narrow range of the electromagnetic spectrum.

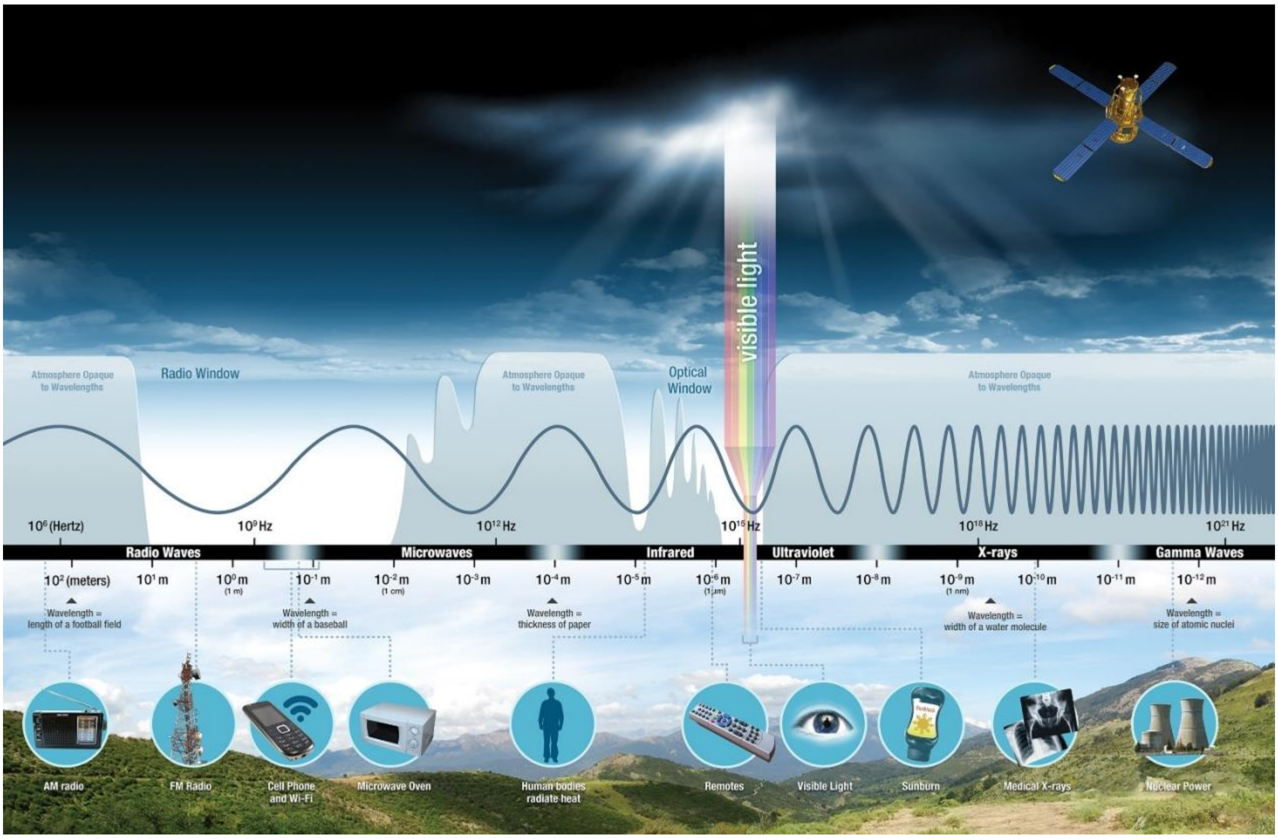


Figure 13. Diagram of the electromagnetic spectrum. The wavelengths used in EO start from the visible and extend to the radio waves, including the infrared and microwaves. Source: NASA Science.

Remote sensing gives humans the superpower to see electromagnetic energy and how it interacts with the matter beyond the visible range to cover the infrared and microwave regions. These spectral regions are subdivided into ranges (e.g., near infrared [NIR], shortwave infrared [SWIR]) and again into bands (e.g., blue, green, red). Everything on Earth reflects, absorbs, and transmits energy. The respective amount reflected electromagnetic radiation as a function of wavelengths defines a unique signature, known as the **spectral signature** of materials. There are spectral libraries that provide the specific spectral information for different materials and that can help identify and differentiate between materials (e.g., minerals, soils, vegetation).

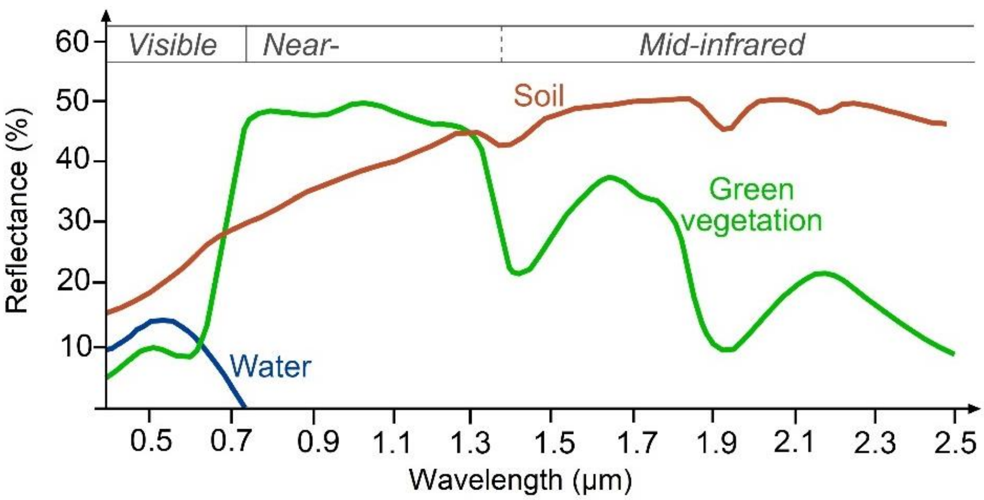


Figure 14. Reflectance of water, soil, and vegetation at different wavelengths. Source: https://seos-project.eu/classification/classification-c01-p05.html.

Another important concept in remote sensing is resolution, which defines the ability to distinguish details in an image and affects how the data can be used. There are four types of resolution: (1) spatial, (2), temporal, (3) spectral, and (4) radiometric.

The **spatial resolution** of an image refers to the size of a pixel in terms of ground dimensions. It provides information on the level of detail that can be captured and the ability to distinguish two closely spaced objects on an image. For square pixels, a 30-meter spatial resolution means that the pixel covers an area of 30m by 30m, or 900 m^2, on the ground. The image resolution depends on the size of the detector and the elevation of the sensor. The smaller the number describing the resolution, the more details you can see.

Figure 15. Landsat 8 image of Reykjavik, Iceland, acquired July 7, 2019, illustrating the difference in pixel resolution. Source: NASA Earth Observatory.

Temporal resolution reflects the revisit time for the same area by a satellite orbiting the Earth—that is, the time interval between when images of the same place on Earth are collected. It depends on the satellite orbit but also on the sensor characteristics and the swath width. Polar-orbiting satellites have 1- to 16-day intervals. MODIS, for example, has a 1- to 2-day temporal resolution, which is ideal for monitoring daily changes. Landsat, with a narrower swath width, has a 16-day temporal resolution, so it can capture changes on a bimonthly basis.

Spectral resolution refers to the wavelength interval or bandwidth that a detector records and thus the ability of a sensor to capture narrower bands of the electromagnetic spectrum. The narrower the range of wavelengths for a given band, the finer the spectral resolution. Multispectral instruments (MSI) capture more than one band and include different intervals of the electromagnetic spectrum (3, 10, 20 bands). Hyperspectral instruments can have hundreds or thousands of bands, like the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), with 224 contiguous spectral channels (bands) with wavelengths from 400 to 2,500 nanometers.

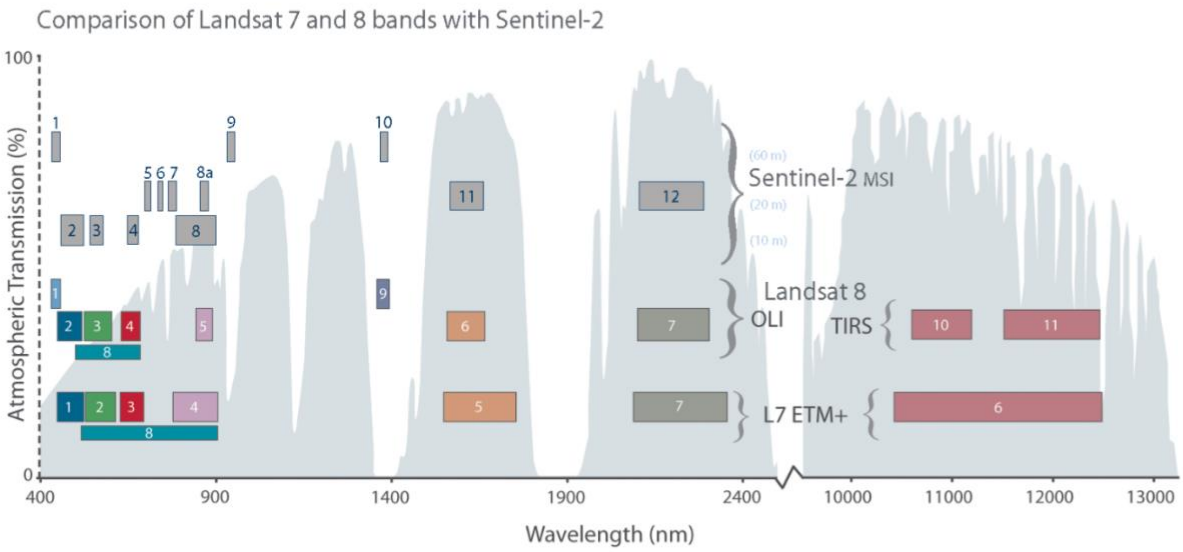


Figure 16. Comparison of Sentinel-2 with Landsat 7 and 8. The specific placement of the Sentinel-2 data is shown above, with very similar spectral bands to Landsat 8 (excluding the thermal bands of Landsat 8's Thermal Infrared Sensor). The visible and near-infrared Sentinel-2 bands have a spatial resolution of 10 meters. The "red-edge" (red and near-infrared bands) and two shortwave infrared bands have a 20-meter spatial resolution. The coastal/aerosol, water vapor, and cirrus bands have a spatial resolution of 60 meters. Source: https://landsat.gsfc.nasa.gov/wp-content/uploads/2015/06/Landsat.v.Sentinel-2.png.

Radiometric resolution, also known as bit depth, is the amount of information stored in each pixel or the brightness levels recorded by a sensor. It is expressed in bits. Radiometric resolution indicates how sensitive an instrument is to small differences in electromagnetic energy. A bit is the smallest unit of data on a computer. It stores only the binary value 0-1. The radiometric resolution is the number of bits representing the energy recorded by the imaging system for a given range of values, and each bit records 2^n radiometric levels. For an 8-bit image, or 2^8, the imaging system can store 256 digital values ranging from 0 to 255 (e.g., Landsat images have 8-bit radiometric resolution). A 16-bit image, or 2^{16}, can store 65,536 digital values or brightness levels. There is trade-off in terms of image size as the bit depth increases.

Annex 3. Case study projects

Table A3.1. Recent impact evaluation projects integrating a remote sensing component into their methodology. The first column (left) indicates which type of variable is the focus of the remote sensing analysis.

	Innovation	Location	Project goals	Remote sensing objectives	Remote sensing data	Geospatial methods
Adoption	Conservation agriculture (CA)	India (Indo-Gangetic Plain)	Provide regionally representative estimates of CA adoption in the IGP region.	Map area of zero tillage and conventional tillage as a measure of CA adoption.	• Sentinel-2 multispectral and Sentinel-1 radar data • Field training data on zero tillage and conventional tillage • Household survey	• Image segmentation using Sentinel-2 to create homogenous zones. • Tillage detection with Sentinel-1, using random forest classifier
Adoption	Faidherbia albida fertilizer tree	Zambia	Evaluate the adoption of the fertilizer tree, Faidherbia albida, agroforestry practice among farmers in Zambia.	Mapping the distribution of the Faidherbia albida fertilizer tree species in Zambia with remote sensing.	• Landsat 8	• Use Landsat 8 satellite imagery and field data to map the presence or absence of the Faidherbia albida fertilizer tree species in agricultural fields
Adoption	Alternate wetting and drying (AWD)	Vietnam	Evaluate the adoption of AWD as a water saving practice in rice production, in the Mekong River Delta of Vietnam.	Determine if remotely sensed data can be used to assess the geographic extent and degree of adoption of the AWD practice. Evaluating the geographic extent and spatial characteristics of areas adopting vs non-adopting, including soil types, location relative to the coast.	• Sentinel-1a and b radar data	• Use Wetness index calculated from Synthetic Aperture Radar (SAR) data from each individual cell, calibrated with soil moisture meters to understand change over time (i.e., patterns of flooding and "dry-down") throughout the growing season to identify AWD fields.

Outcomes	Index-based livestock insurance (IBLI)	Kenya and Ethiopia	Evaluate the environmental impacts of index-based livestock insurance using remotely sensed data to capture spatiotemporal variation in rangeland health and look at the causal impacts of IBLI adoption on rangeland health.	Evaluation the environmental externalities of IBLI at scale in Kenya and Ethiopia between 2010 and 2019 by developing a new rangeland health (RH) index • The RH index accounts for soil/site stability, hydrologic function and biotic integrity, measured through the presence of bare ground, site potential deviations and Solar induced chlorophyll fluorescence (SIF) and leaf area index (LAI).	• Ground-based geo-tagged photos for ground truthing and ground-based measurement; • Very high-resolution imagery (Worldview I, II, III, Geoeye-1) • Landsat 5, 7 and 8, MODIS and OCO-2 sensors	Develop RH assessment indicators of RH attributes to create a new RH index using direct and indirect remote sensing estimates, including: • Identification of bare ground and functional group canopy cover through machine-learning based on sub-pixel classification of fractional cover of very high and mid-resolution imagery; • Assess biotic integrity (e.g., annual production) using SIF and LAI time series, including to measure site potential deviations.
Outcomes	Restoration of the commons	India	Evaluating the impacts (including the longer-term ecological and socioeconomic impacts) of a large-scale land restoration initiative in India as well as the model of intervention of Foundation for Ecological Security.	Remote sensing is used to evaluate the presence of a measurable restoration of common lands, including associated enhancement in ecosystem services such as carbon sequestration and to understand spatial variation in ecological outcomes.	• Combination of field inventory and remote sensing data from a range of different platforms (e.g., MODIS, Landsat, RapidEye and Sentinel-2)	• Ecological health assessment using the Land Degradation Surveillance Framework used for identification of land degradation hotspots at the landscape level, that rely on field and remote sensing measurements of vegetative cover and soil characteristics (carbon, erosion, infiltration capacity).

Outcomes	Sorghum and millet upscaling project	Mali	Evaluate the impacts of ICRISAT's sorghum and millet interventions on land-use and land-cover change (LULCC) for the Sikasso region of Mali, including the intensification induced by the technology improvement of the SMS intervention and cotton cash cropping.	Geospatial impact evaluation to address environmental externalities related to LULCC relying on three remotely sensed outcome metrics (indicators) to measure LULCC in the region from 2002 to 2020: • Tree cover, cropland extent, and landscape composition and configuration.	• Landsat at 30m resolution • 2019 tree canopy map at 0.5m resolution produced (Brandt et al., 2020) • FAO Collect Earth for training data and field data collection for validation data	• Wall-to-wall mapping of cropland extent • Percent tree cover density within and outside croplands obtained from remote sensing, tree canopy map and Collect Earth platform. • Landscape composition and configuration through analysis of LULC classes, landscape indices of the different classes of tree cover density and possible automated algorithm built on the FAO training dataset.
Outcomes	Direct seed marketing program	Ethiopia	Evaluate if the Direct Seed Marketing (DSM) pilot program, which is a partial liberalization of the seed market in Ethiopia since 2011, has increased agricultural productivity through better access to improved seed and if it has improved the timing of planting following the start of the rainy season.	Focusing on maize, this study directly evaluates the role of remote sensing for studying impacts of this type of program. It focuses on three potential impacts: • Agricultural productivity • Timing of planting following the rainy season onset. • Length of growing season.	• MODIS 8-day vegetation index at 250m (MOD13Q1 and MYD13Q1) • Copernicus Global Land Service (CGLS) • CHIRPS rainfall data set • Gridded rainfall onset product • Elevation (worldclim) and accessibility to cities (Weiss et al. 2018)	• Phenological analysis based on a time series of fitted value of harmonic model on EVI vegetation indices from MODIS 250m resolution for the period 2002 to 2019 extracted at the woreda (district), kebele and pixel levels • Mapping of maize/non-maize agricultural land for the time series • Extraction of covariates for causal inference model testing
Both adoption and outcomes	Demi-lune (DL) rainwater harvesting technique	Niger	Measure the medium-term impacts of DL, implemented in 2018 as part of a Randomized Control trial (RCT) providing training and financial incentives. Looking at the sustained and new adoption, land degradation, abandonment and extensification, and environmental spillovers.	Use remote sensing to monitor adoption, dis-adoption, and spillover of DL: • Count DL or area covered by DL in the project area. • Measure soil moisture changes at the pixel level. • Measure crop production at the pixel level.	• Very high- resolution imagery (<1m) • PlanetScope • Sentinel-2	• Delimit DL field perimeter using very high-resolution imagery • Use the field perimeters and vegetation/soil moisture indices derived from PlanetScope and/or Sentinel-2 to assess difference in soil moisture and crop productivity between DL fields and control fields.

Both adoption and outcomes	Happy seeder (direct-seeders for wheat sowing)	India (Indo-Gangetic Plain)	Evaluate the impacts of Happy seeder diffusion, a technology that allows to sow wheat under the rice residues, on residue burning and associated air pollution. Burning of rice residues are widespread under intensive, irrigated rice-wheat systems of the northwestern Indo-Gangetic Plains and the Happy seeder reduces the need for burning.	Evaluate the environmental spillover related to evidence of burning using: • Fire detection by satellite • Detection of emission of gases from biomass burning by satellites Triangulated by village surveys and farm-household surveys	• High-resolution Google Earth imagery • Sentinel-2 data (Level-1 C) • Active village-level fire detected by Sentinel-2 and MODIS satellite instruments • GHG measurements from Sentinel-5P (TROPOMI) and OCO-2	• Mapping zero tillage and conventional tillage using Sentinel-2 spectral bands and derived VI for classification • Locations of fire affected soundings (extent and intensity) from Terra and Aqua satellite observations and Sentinel-2 • Sentinel-5 TROPOMI and atmospheric transport model to determine GHG enhancement relative to background biomass burning, decoupling emissions of crop residue burning from other sources.
Both adoption and outcomes	Stress-tolerant rice varieties (STRVs)	Bangladesh	Evaluate the adoption and impact at scale of submergence-tolerant rice varieties. The study includes a mesoscale analysis based on remote sensing and a microscale analysis of income, consumption, and food security using household data.	Provide evidence of STRV adoption by measuring the impacts on post-flood vegetation greening derived from remotely-sensed vegetation indices, in comparison with locations where STRVs were not adopted.	• Google Earth imagery and 500m MODIS, and Landsat (phase 2) • NASA/Darmouth Flood Observatory 250m MODIS Near real time (NRT) Global Flood Mapping product and Tellman et al. (2021) (phase 2) • MOD09A1 MODIS/TERRA Surface Reflectance 8-Day 500m and Landsat (phase 2)	• Mapping rice crop areas using MODIS 500m/250m and GE imagery for training and validation; • Mapping flooded area and flooded rice areas using a flood mapping product and calculating the flood incidence and flood intensity index; • Per-pixel phenology based on the Start of Season and Peak of Season Enhanced Vegetation index (EVI) as well as the Land Surface Water Index (LSWI).

Both adoption and outcomes	Improved forages	Ethiopia	Evaluate the impacts of the adoption of improved forages on livelihoods and the environmental spillover on soil erosion and tree cover, as well as identify the institutional factors promoting the adoption and diffusion of this technology (scaling).	Spillover of adoption with RS, including: • Measure of land use, biomass production and tree cover change • Soil moisture, soil erosion	• Sentinel-2 and PlanetScope data • GPS and other field measurements for model building	• Land use classification of the improved forage area using Sentinel-2 and modelling of biomass productivity with change in NDVI • Compare field, modelled and remote sensing measurements of soil moisture and soil erosion

Annex 4. Lessons learned from case studies

Table A4.1. Key lessons learned from case studies about using remote sensing for impact evaluation

Case study	Key lessons learned
Conservation agriculture (CA), India (Indo-Gangetic Plain)	- Importance of randomly selected reference data. One key challenge was that the available ground reference data were not selected randomly; GPS points were located near IRRI experimental sites, with more data points in Punjab and Haryana and fewer in Bihar. There are differences between agricultural systems in Punjab/Haryana and Bihar. The remote sensing analysis results showed higher area estimates of land under zero tillage in Bihar than expected. - Self-reported zero tillage adoption collected through the survey did not contain visual aids, and the characteristics of zero tillage were not necessarily clear to people, leading to possible response bias and affecting area estimates. The survey results were planned to be used to compare the accuracy of EO-based estimates. - Faced with a discrepancy between the EO-based and survey-based estimates, it was not clear which estimate was sensibly more accurate because of methodological issues in both and/or possibly because of partial zero tillage adoption. The project would have needed additional ground reference data to perform a proper accuracy assessment of the ZT/CT map and to improve the reliability of the results. The challenge is that the data sources do not complement each because of errors affecting each source.
Faidherbia albida fertilizer tree, Zambia	- The relatively low level of adoption of the fertilizer tree in Zambia (6% among survey sites and 15% of cultivated fields surveyed) signifies that the fertilizer tree occupies a small area of the landscape and that less training data may be available, leading to lower accuracy in predicting the presence of F. albida (60%) than its absence (93%) (Stevenson and Vlek, 2018). - Farmers may leave fertilizer trees when clearing woodlands for cultivation, which means that the adoption pathway is not clear. They may also retain other "useful" tree species on their fields, which may reduce the mapping accuracy of Faidherbia albida.
Alternate wetting and drying (AWD), Vietnam	- This analysis covered only one rice-growing season (three months) and used Sentinel-1 A and B with a six-day repeat cycle for the constellation, but the accuracy of AWD detection may have been improved by covering more than one crop season. - Being able to collect additional repeated measurements of the water depth level would have helped better account for random errors in training data and for the within-field variability in water level, thus improving the calibration of Sentinel-1 imagery. - There is high variability in the landscape between farmers, regions, and soil types, which may have affected the accuracy of detection needed for mapping adoption of AWD and assessing accuracy. - Being able to triangulate or complement information obtained from a survey with remote sensing, or zooming in to some area with precise information about the practice, would have helped improve confidence and robustness.

Index-based livestock insurance (IBLI), Kenya and Ethiopia	- The large intervention area affects the feasibility of collecting geographically representative reference data. Geotagged photographs of land cover types can be used productively as a substitute for ground truthing data for past conditions or when field data cannot be collected. Exploring the potential of citizen science to support data collection at scale, even if errors may be introduced, is an important avenue to investigate for future geospatial impact evaluations taking place over a large area. - The access by the project to a large volume of very high-resolution imagery would have required the support of a computer scientist for processing of these VHR images, which needed to be pre-processed before analysis—a huge burden. This pre-processing step has taken much more time than anticipated. - Determining the adequate scale or range of the generating process can be a challenge in itself. The project includes information about sales of insurance policies by index administrative units and how they rolled out over space and time, but it doesn't consider the migration of livestock in the landscape. For this reason, the research team had to integrate additional information from migration patterns detected by telemetry, using a collar that tracks animal movement, to capture the scale of the impact of the migrating livestock on rangeland health.
Restoration of the commons	- Organizing the field campaign for this type of measurement is a challenge. The project faced delays due to the global pandemic, which impeded international shipping and raised bureaucratic hurdles for sending soil samples from India to Kenya. There were other delays in soil laboratory analysis because of equipment failure, impacting access to reference data for the remote sensing analysis. - Because of the established Land Degradation Surveillance Framework, the project only had to sample the sites at the end of the land restoration project to calibrate the model to local conditions. The change in ecosystem services is evaluated through changes in EO-based indicators. This demonstrates the benefit of having a modeling framework with consistent methodology replicated over many countries. - The time lag for the agroforestry or tree planting project is substantial, so a geospatial impact evaluation needs to be structured to account for the time lag before measuring impacts, including building sophisticated models to reconcile the time series of satellite data for a long time period.
Sorghum and millet upscaling project, Mali	- It can be a challenge to explain the need for remote sensing analysis or validation of image interpretation to on-the-ground partner organizations if they do not have a background in remote sensing. - When using existing products (e.g., tree cover for Africa at 1-m resolution), it is important to understand the error structure of the products and the data-generating process. - There is important variation in mapping accuracy between different landscapes; robust remote sensing analysis should be validated with ground verification.
Direct Seed Marketing program, Ethiopia	- Requesting reference data from national household survey should be done at the onset of the project because delays in accessing these data can delay the project. - Using existing household panel survey results as a benchmark to develop a geospatial impact evaluation provides strong advantages for this complex remote sensing workflow. Without this previous study, which showed significant impacts of DSM on maize yields, the project would have lacked the appropriate pathway to follow because of all the possible confounding

	factors, including how the program was implemented. - The lack of available geospatial products (Essential Agricultural Variables) can be a major challenge for a geospatial impact project, creating the need to generate all of the spatial inputs or to rely on coarse assumptions.
Demi-lune (DL) rainwater harvesting technique, Niger	- Take full advantage of prospective analysis to plan for data access, field data collection, and piloting/testing approaches. - The costs of very high-resolution imagery can be prohibitive for large areas without an agreement with a provider, affecting the options for imagery. - The required reference data may differ between a project's objectives; for adoption, point data may be sufficient, but for measuring impacts, collecting plot boundaries (perimeters) of the treated and control fields may be more suitable.
Happy Seeder (direct seeders for wheat sowing), India (Indo-Gangetic Plain)	- In case of illegal crop residue burning, farmers will underreport their participation in surveys, so the survey data is expected not to be usable for comparison of EO-based estimates. The EO-based estimates should thus have rigorous independent accuracy assessment to demonstrate validity. - Many factors may impact geospatial impact evaluation, even outside of the remote sensing realm, but still influence the conditions for evaluation with remote sensing methods. The project faced challenges with farmers switching from Happy Seeder to other direct seeders; conflict between farmers and government, which reduced interest in participation in research; and the large-scale abandonment of farming as a whole by farmers. - Combining researchers from very different fields can provide an innovative project linking technology adoption, conservation agriculture, crop residue burning, and greenhouse gas emissions with high societal relevance.
Stress-tolerant rice varieties (STRVs), Bangladesh	- Older products, such as the DFO flood maps, can be less comprehensive than more recent products and need to be examined (metadata) and validated with more accurate products (e.g., Sentinel-1 flood maps). - The lack of available reference data on field boundaries for training and validation (with and without STRVs), both in present time and in the past, can be a major impediment to more accurate maps and therefore impact results. - The resistance of organizations to collect georeferenced data in the field requires a cultural change so that remote sensing can be leveraged at its full potential. - Real interdisciplinary work in geospatial impact evaluation requires researchers to devote time and effort to achieve mutual understanding, but it also contributes to pioneering insights.
Improved forages, Ethiopia	- For mapping improved forages adoption, the area of intervention is very small compared with other land covers; thus, it is challenging to collect sufficient area (pixels) of improved forages for training the machine-learning algorithm and to obtain a balanced sample. This had to be resolved through bootstrapping selection of pixels under improved forages. The fact that another land cover, grassland, is similar to the improved forages also created a challenge for classification. - Variation over the Ethiopian highlands landscape presents a challenge for geographic representativeness, requiring more reference data for accurate classification.

Annex 5. EO data providers, data sets and tools

This section provides a non-exhaustive list of free and open data EO providers where readers can access and download EO data or directly perform their analysis.

Table A5.1. Data providers, tools and services to support access and use of EO data

Name	weblink	Short description
CEOS Data Cube	https://www.opendatacube.org/ceos	Initiative to support the development of International Data Cube to facilitate access to ARD to global users
Earth System Data Cube (ESDC)	http://earthsystemdatacube.net	ESDL offers a framework to effectively map user-defined functions (UDFs) to these data cube with access to highly-curated analysis ready data. The interface support python, Julia and is in development for R.
Google Earth Engine	https://earthengine.google.com/	Petabits of data from the catalogue, with both satellite data and derived products
Open Data Cube	https://www.opendatacube.org/	Open-source platform for accessing, managing, and analyzing EO data with python
DIAS, European Copernicus Programme	https://www.copernicus.eu/en/access-data/dias	Data and Information Access Services includes five platforms funded by the European Commission that gives to Sentinel and commercial imagery, as well as other services.
Sentinel Hub	https://www.sentinel-hub.com https://apps.sentinel-hub.com/eo-browser/	Official headquarters to download Sentinel imagery
Copernicus Sentinels Open Access Hub	https://scihub.copernicus.eu/dhus/#/home	Copernicus data access portal
Coastal Thematic Exploitation Platform	https://www.coastal-tep.eu	Platform of data services focused on coastal areas
Forestry Thematic Exploitation Platform	https://f-tep.com	Platform of data services focused on global forests
Hydrology Thematic Exploitation Platform	https://hydrology-tep.eu	Platform of data services focused on hydrology
Geohazards Thematic Exploitation Platform	https://geohazards-tep.eu	Platform of data services focused on disasters and risk management
Urban Thematic Exploitation Platform	https://urban-tep.eu	Platform of data services focused on urban areas

Food Security Thematic Exploitation Platform	https://foodsecurity-tep.net	Platform of data services focused on agriculture and food security
Trends.Earth	http://trends.earth	Open source tool to look at land change
NextGEOSS	https://nextgeoss.eu	European data hub and platform, a project with focus on applications for businesses but with potential relevance for scientific applications
Global Earth Observation System of Systems (GEOSS)	https://www.geoportal.org/	Global data portal and access point for users seeking data, imagery and analytical software packages relevant to all parts of the globe
GEO BON	https://geobon.org https://portal.geobon.org/home	Essential Biodiversity Variables data portal, bringing data from different sources including geospatial data
GEOGLAM	http://geoglam.org	Group on Earth Observations Global Agricultural Monitoring Initiative (GEOGLAM) provides information about products relevant to global, regional and national agricultural monitoring with EO; CGIAR is part of GEOGLAM organization
GEOGLAM - Crop monitor	https://cropmonitor.org/	Global platform for EO data and tools
Global Information and Early Warning System on Food and Agriculture (GIEWS)	https://www.fao.org/giews/earthobservation/index.jsp?lang=en	FAO EO platform for agriculture monitoring
Famine Early Warning System (FEWS NET)	https://fews.net/	FEWS NET Data Center provides access to different data portal relevant to agriculture monitoring
Anomaly Hotspots of Agricultural Production	https://mars.jrc.ec.europa.eu/asap/	ASAP provides EO and other products relevant to agriculture monitoring
GEOGLAM RAPP	https://map.geo-rapp.org/	Geospatial tool relevant to Rangeland and pasture Productivity (RAPP)
GEO Blue Planet	https://geoblueplanet.org	EO-based variables relevant to ocean and coastal observations (under development)
Global Forest Observation Initiative (GFOI)	http://www.fao.org/gfoi	Data coordination will support the acquisition, availability and accessibility of remote sensing data and other datasets and tools for forest monitoring (under development)
GEO-Wetlands	https://geowetlands.org	The Group on Earth Observations Global network for Observations of wetlands
GEO Global Network for Observation and Information in Mountain Environments	https://mountainresearchinitiative.org/activities/projects/geo-mountains	The Group on Earth Observations Global Network for Observations and Information in Mountain Environments.

CEOS Ad-Hoc Team on SDGs	www.ceos.org/sdg	Provide a list of data providers relevant to SDGs. The page is in development at the time of writing but should become a reference in the future.
Global Forest Watch	https://www.globalforestwatch.org	Platform that provides EO data and tools for monitoring forests.
Global Mangrove Watch	https://www.globalmangrovewatch.org	Platform that provides EO data and tools for monitoring mangrove forests.
Global Human Settlement Layer	https://ghsl.jrc.ec.europa.eu	Data and tools for assessing the human presence on the planet
Global Surface Water Explorer	https://global-surface-water.appspot.com/download	Data product of high-resolution mapping of global surface water and its long-term changes
Freshwater Ecosystems Explorer	https://www.sdg661.app	Data platform for freshwater ecosystems relevant to SDG6
ESA Climate Change Initiative Land Cover	https://climate.esa.int/en/projects/land-cover	Provides updated land cover related products
Global Soil Organic Carbon map	http://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/global-soilorganic-carbon-map-gsocmap	Soil Maps and Soils Databases
International Soil Reference and Information Centre	https://www.isric.org/explore/soilgrids https://soilgrids.org/	ISRIC soil data provider
Land Processes Distributed Active Archive Center (LP DAAC)	https://lpdaac.usgs.gov/	Data Center for Land Processes DAAC, of NASA EOSDIS
USGS EarthExplorer	https://earthexplorer.usgs.gov/	Provides access to Landsat data, MODIS et much more.
NASA Earthdata Search	https://search.earthdata.nasa.gov/search	Provides access to all NASA DAAC data
Application for Extracting and Exploring Analysis Ready Samples (AppEEARS)	https://appeears.earthdatacloud.nasa.gov/	An application to access and transform geospatial data from a variety of US federal data archives
Oak Ridge National Laboratory Distributed Active Archive Center	https://daac.ornl.gov/	Data Center for Biogeochemical Dynamics, a NASA Earth Observing System Data and Information System (EOSDIS) data center
Socioeconomic Data and Applications Center (SEDAC)	https://sedac.ciesin.columbia.edu/	Data Center for socioeconomic data, a NASA Earth Observing System Data and Information System (EOSDIS) Center
WorldClim - Global Climate Data	http://www.worldclim.com/version2	Global climate data for modeling and GIS
FAO Land & Water databases	https://www.fao.org/land-water/databases-and-software/en/	

NASA SERVIR catalogue	https://gis1.servirglobal.net/geonetwork/srv/eng/catalog.search#/home	NASA SERVIR program has its own data catalogue of global/regional products
Digital Earth Africa	https://www.digitalearthafrica.org/	Regional platform for Africa
ESRI Africa Geoportal	https://www.africageoportal.com/	Open GIS/RS data for Africa
openAFRICA	https://africaopendata.org/	Open GIS/RS data for Africa
World Meterological Organization - OsCAR	https://space.oscar.wmo.int/	Global repository of surface and space observing system capabilities
NOAA Data Access Viewer	https://coast.noaa.gov/dataviewer/#/	National Oceanographic and Atmospheric dataview to download authoritative land cover, imagery, and lidar data.
DigitalGlobe Open Data Program	https://www.maxar.com/open-data	DigitalGlobe's Open Data Program supplies satellite imagery for relief for any natural disaster,
NASA Worldview	https://worldview.earthdata.nasa.gov/	Data provider for download of scientific EO products
NOAA Comprehensive Large Array-data Stewardship System (CLASS)	https://www.avl.class.noaa.gov/saa/products/welcome	Data provider for atmospheric, climatic, and environmental EO data
JAXA mission data	www.eorc.jaxa.jp/en	JAXA free and open datasets
EUMETSAT mission data	https://navigator.eumetsat.int/start	Data provider for European operational satellite agency for monitoring weather, climate and the environment
National Institute for Space Research (INPE)	http://www.dgi.inpe.br/	Data provider especially for CBERS (China-Brazil joint mission), data is specific to South America and Africa
	http://www2.dgi.inpe.br/catalogo/explore	
DIVA-GIS	https://www.diva-gis.org/Data	Free GIS data by country; developed by Robert Hijmans et al., supported by CGIAR.
Bhuvan Indian Geo-Platform of ISRO	https://bhuvan-app3.nrsc.gov.in/data/download/index.php	Indian Space Research Organisation geoportal; strong India focus

Annex 6. Global indicators related to SDGs and One CGIAR strategic goals

In this section, we showcase the results of an assessment of the overlap between satellite-based SDG indicators and the One CGIAR five impact areas. We highlight findings by O'Connor et al. (2020) that identified 34 indicators that can be measured directly or indirectly with satellite data and evaluated their readiness and adequacy. For readiness, they looked at the maturity of EO technologies, the status of EO indicator guidelines, the technical capacity required, and the availability of global EO data. To evaluate adequacy, they look at sensitivity to change, the scalability of the existing EO approach, and other factors. In the Table A6.1, we show the score in green for a high level of relevance and robustness of EO data and in yellow when EO should be complemented by or complementing other data sources.

Table A6.1. Mapping of global indicators relevant to SDGs and One CGIAR strategic goals

One CGIAR impact area	Related SDGs	SDG indicator description	Score	Data sources (with weblink)
Poverty reduction, livelihoods, and jobs	**SDG 1: Zero Poverty**	1.1.1 Proportion of population below the international poverty line, by sex, age, employment status and geographical location (urban/rural)		WorldView, GeoEye, QuickBird, IKONOS satellite imagery
		1.2.1 Proportion of population living below the national poverty line, by sex and age		Landsat satellites
		1.4.1 Proportion of population living in households with access to basic service		DSMP, VIIRS, and harmonized product
		1.5.2 Direct economic loss attributed to disasters in relation to global gross domestic product (GDP)		The Global Urban Footprint (GUF) / World Settlement Footprint (WSF)
				The Global Human Settlement Layer (GHSL)
				Emergency Events Database (EM-DAT)
				Disasters and conflicts: UNEP Data Explorer
Nutrition, health, and food security	**SDG 2: Zero Hunger**	2.3.1 Volume of production per labor unit by classes of farming/ pastoral/forestry enterprise size		Copernicus Dry matter productivity product
		2.4.1. Proportion of agricultural area under productive and sustainable agriculture		GEO Global Agricultural Monitoring (GEOGLAM)
				Sentinel-2 for agriculture monitoring

				Sentinels for Common Agriculture Policy
				Sentinel-2
	SDG 3: Good Health and Well-Being	3.9.1 Mortality rate attributed to household and ambient air pollution		Satellite Aura
				Sentinel-5P
				MODIS/VIIRS Air quality
				MISR
	SDG 11: Sustainable Cities and Communities	11.6.2 Annual mean levels of fine particulate matter (e.g., 2.5 and PM10) in cities		Gases Observing Satellite (GOSAT)
				Carbon Observatory-2 (OCO-2)
				MODIS Aerosol Product
	SDG 6: Clean Water and Sanitation	6.1.1 Proportion of population using safely managed drinking water services		GlobWetland II
				The Global Human Settlement Layer (GHSL)
				The Global Surface Water Explorer
		6.3.1 Proportion of wastewater safely treated 6.3.2 Proportion of bodies of water with good ambient water quality		The Global Surface Water Explorer
				Bio-physical parameters can be derived from Sentinel-2 and Sentinel-3 OLCI using processors in the SNAP toolbox provided by ESA's Scientific Toolbox Exploration Platform (STEP)
				CyanoLakes
		6.4.1 Change in water-use efficiency over time 6.4.2 Level of water stress: freshwater withdrawal as a proportion of available freshwater resources		Sentinel-1 and Sentinel-2
				MODIS
				ESA CCI Land Cover

				Global Map of Irrigation Areas (GMIA) of FAO
				Global Irrigated Area Map (GIAM) of IWMI
				The FAO portal to monitor Water Productivity through Open access of Remotely sensed derived data (WaPOR)
				Dry matter productivity (yield) and water bodies map from Copernicus Land Services
Environmental Health and Biodiversity	**SDG 15: Life on Land**	15.1.1 Forest area as a proportion of total land area		Landsat data
		15.1.2 Proportion of important sites for terrestrial and freshwater biodiversity that are covered by protected areas, by ecosystem type		Sentinel from the Copernicus Open Access Hub
		15.2.1 Progress toward sustainable forest management		Copernicus Global Land Cover
		a) Forest area net change rate		Global Forest Watch (global change 2000–2021)
		b) Above-ground biomass stock in forest		Pan-tropical biomass map
		c) Forest area within legally established protected areas		Global terrestrial biomass map and ORNL-DAAC above and belowground biomass data, both for 2010
		d) Forest area under a management plan		ESA CCI Land Cover
		e) Forest area under management certification scheme		Global Ecosystem Dynamics Investigation
		15.3.1: Proportion of land that is degraded over total land area, with three sub-indicators		SEEA-MODIS
		a) Trends in land cover		World Atlas of Desertification
		b) Trends in land productivity		MODIS archive
		c) Trends in carbon stocks, above and below ground.		SoilGrids 250m - soil properties global iSDA 30m maps of soil properties for Africa

		15.4.1 Coverage by protected areas of important sites for mountain biodiversity		The World Database on Protected areas, as accessible via Protected Planet
				The World Database on Key Biodiversity Areas, as accessible via BirdLife
				Global Mountain Explorer
				Socio-Economic Data and Application Center
		15.4.2 Mountain Green Cover Index		Shuttle Radar Topography Mission (SRTM) 1 arc-sec global
				ASTER DEM 30 m global
				Precise Global Digital 3D Map "ALOS World 3D"
				OpenTopography
	SDG 14: Life Below Water	14.1.1 Index of Coastal Eutrophication and floating plastic debris density		Landsat 8, 9
		a) Index of Coastal Eutrophication		Sentinel data (1,2, and 3) from the Copernicus Open Access Hub
		b) Surface water chlorophyll		ASTER-Terra
				MODIS-Aqua & Terra
				SeaWiFS
				CZCS
	SDG 6: Clean Water and Sanitation	6.6.1 Change in the extent of water-related ecosystems over time		Landsat
		a) Spatial extent dynamics		Sentinel data (1,2, and 3) from the Copernicus Open Access Hub
		b) Water quality (lakes/ artificial water bodies)		L-Band SAR satellite series operated by JAXA: JERS-1 SAR (1992-1998), ALOS PALSAR (2006-2011), ALOS-2 PALSAR-2 (2014-present)
		c) Water quantity		Terra/Aqua MODIS
		d) Groundwater quantity in aquifers		Visible Infrared Imaging Radiometer Suite (VIIRS) weather monitoring system
				The Global Surface Water Explorer

				FLO1K, a consistent streamflow dataset at a resolution of 30 arc seconds (~1km) and global coverage
				CMAP (CPC Merged Analysis of Precipitation) refers to a collection of precipitation data sets.
				GCP (Global Precipitation Climatology Project)
				GMW (Global Mangrove Watch), consistent dataset of mangrove extent.
Climate adaptation and mitigation	**SDG 7: Affordable and Clean Energy**	7.1.1 Proportion of population with access to electricity		NASA's Black Marble night-time lights product suite (VNP46)
				NOAA VIIRS Night Lights Annual Composites
				POWER
	SDG 13: Climate Action	13.1.1 Number of deaths, missing persons and directly affected persons attributed to disasters per 100,000 population		NASA Hazards and disasters
				Global MODIS Flood Mapping initiative
				Group on Earth Observations Global Agricultural Monitoring
				USGS Famine Early Warning Systems Network
				The ESA Thematic Exploitation Platform (TEP) on Geohazards

Annex 7. Satellite sensors and their characteristics

This section provides an overview of different satellites/sensors and their characteristics. The same satellite dataset can be accessed from various data providers. Annex 5 gives a list of data providers so that researchers can learn where to access data. This annex shows the most common satellite missions so that researchers can identify what they are looking for based on their specific goals.

Table A7.1. Summary of the most common satellites/sensors and their characteristics

Sensors/Satellites	Wavelength	Spatial resolution (m)	No. of bands /Polarization	Swath width (km)	Repeat cycle (days)	Lifetime	Access	Agency	Purpose
Optical: Coarse resolution									
AVHRR	0.580-12.5 µm	1,000	6	2,900	daily	1998-	Free	NOAA	Multi-purpose imagery
MODIS	645 nm-14.235 µm	250, 500, 1,000	36	2,330	daily	2002-	Free	NASA	Multi-purpose imagery
VIIRS	412 nm-11.45 µm	400/750	22	3,000	daily	2011	Free	NOAA/NASA	Collects images and radiometric data used to provide information on the Earth’s clouds, atmosphere, oceans, and land surfaces
Optical: Mid-resolution									
Landsat-5 TM	0.45-12.5 µm	30	7	185	16		Free	USGS/NASA	Land Observation; Thematic Mapper
Landsat-7 ETM+	0.45-12.5 µm	30	8	185	16		Free	USGS/NASA	Land Observation; Enhanced Thematic Mapper +
Landsat-8 OLI	433-2,300 nm	30	9 (+ 2 TIR)	185	16	2013	Free	USGS/NASA	OLI instrument and Thermal Infrared Sensor (TIRS) with 2 channels at 100m spatial resolution
Landsat-9	433-2,300 nm	30	9 (+ 2 TIR)	185	16	2021	Free	USGS/NASA	Land Observation; OLI-2 and TIRS-2 instruments
ASTER	0.52-11.65 µm	15, 30, 90	15	60	16	1999	Free	METI	High-resolution land and vegetation observation

Sensors/Satellites	Wavelength	Spatial resolution (m)	No. of bands /Polarization	Swath width (km)	Repeat cycle (days)	Lifetime	Access	Agency	Purpose
ALI	433-2,350 nm	30	10	37	16	2001-2017	Free	NASA	Advanced technology for high-resolution land and vegetation observation
SPOT 1-5	0.545-1.64 µm	2.5-20	15	60	3-5	1986-2015	Scientific research and application	CNES	High-resolution land and vegetation observation. Digital Elevation Model (DEM)
Sentinel-2 MSI	443-2,190 nm	10, 20, 60	13	290	Satellite: 10 Constellation: 5	2015-	Free	ESA	High-resolution land observation, vegetation, territory management, and hazards mitigation
Sentinel-3	400-1,020 nm	300	21	~1270	hourly to montly	2016	Free	ESA	Ocean and land observation. OLCI: Ocean and Land Colour Image, e.g. Colour Dissolved Organic Matter (CDOM), Ocean chlorophyll concentration.
IRS Infrared scanner	8-12 µm	73	1	18		2013-		CAST	High resolution land observation; TIR radiometer
Very high resolution									
SPOT 6-7	0.45-0.89 µm	1.5 (PAN); Multispectral: 6	4	60	daily	2012-	Scientific research and application	Airbus Defence and Space	New Astrosat Optical Modular Instrument (NAOMI) and VEGETATION instruments
Pleiades-1A	0.49-0.83 µm	0.7 (PAN) Multispectral: 2.8	4	20	Satellite: 26 Constellation: 2	2011-		CNES	Very-high-resolution land and vegetation observation. Digital Elevation Model (DEM), with HiRI instrument High-Resolution Imager
GeoEye	450-900 nm	0.41 (PAN) Multispectral: 1.64	5	15.2	3	2009-	Commercial; Partial	GeoEye (Maxar)	Very-high-resolution land imagery
IKONOS	450-900 nm	1-4	4	11.3	5	1999-2015	Commercial; Partial	GeoEye (Maxar)	Very-high-resolution land imagery; Ball Global Imaging System 2000

QuickBird	450-900 nm	0.60 (PAN) Multispectral: 2.4	4	18	5	2001-2015	Commercial; Partial	DigitalGlobal	
Sensors/Satellites	**Wavelength**	**Spatial resolution (m)**	**No. of bands /Polarization**	**Swath width (km)**	**Repeat cycle (days)**	**Lifetime**	**Access**	**Agency**	**Purpose**
WorldView 2-3 (WV110)	400-1,040 nm	0.46 (PAN) Multispectral: 1.84	9	17.7	~3	WV2: 2009 WV3: 2014	Commercial; Partial	DigitalGlobe	Very-high-resolution land imagery
PlanetScope	485-820 nm	3.7	4	24.6	Daily	2014-	Partial	Planet	High-resolution land observation and disaster monitoring; CubeSat: Constellation of nano-satellites; FLOCK instrument
RapidEye	440-850 nm	6.5	5	78	Daily	2008-2020		DLR	Very-high resolution land observation and disasters monitoring; 5 satellites
Skysat	450-900 nm	0.9 (PAN) Multispectral: 2.0	5	8	Daily	2014-		Terra Bella	Very-high-resolution land imagery
RSI (FORMOSAT-5)	0.485-0.830 µm	2 (PAN); Multispectral: 4	5	24	daily	2017-		NSPO , UCAR	High-resolution land observation for vegetation and disasters monitoring; FORMOSAT-2 (2004-2016)
KOMPSAT-3A	450-900 nm	0.7 (PAN) Multispectral: 2.8	5	15	~3	2012-		KARI	High-resolution land and vegetation observation, DEM; AEISS instrument
Synthetic Aperture Radar									
ERS-1 (AMI-SAR)	5.6 cm (C-band)	Az: 6-30 Rg: 26	VV	100	35	1991-2001	Partial	ESA	High-resolution all-weather multi-purpose imager for ocean, land and ice. Also wave spectra.
JERS-1 (SAR)	24.6 cm (L-band)	Az: 18 Rg: 18	HH	75	44	1995-1998	Partial	JAXA	Survey of geological phenomena, land usage (agriculture, forestry), observation of coastal regions, geologic maps, environment, disaster monitoring, etc.

Sensors/Satellites	Wavelength	Spatial resolution (m)	No. of bands /Polarization	Swath width (km)	Repeat cycle (days)	Lifetime	Access	Agency	Purpose
ERS-2 (AMI-SAR)	5.6 cm (C-band)	Az: 6-30 Rg: 26	VV	100	35	1995-2011	Partial	ESA	High-resolution all-weather multi-purpose imager for ocean, land and ice. Also wave spectra.
RADARSAT-1	5.6 cm (C-band)	Standard: 25X28 Fine: 9X9 Wide1: 35X28 Wide2: 35X28 ScanSAR: 50X50	HH	Standard: 100 Fine: 45 Wide1: 165 Wide2: 150 ScanSAR: 305-510	24	1995-2013	Commercial	CSA	Multi-purpose SAR observation, especially for ice.
ENVISAT (ASAR, AATSR)	5.6 cm (C-Band)	Az: 28 Rg: 28	HH,VV, VV/HH HH/HV, VV/VH	100	35	2002-2012	Partial	ESA	Atmospheric chemistry, climatology, ocean and ice. ENVISAT also has the SCIAMACHY instrument.
ALOS (PALSAR)	24.6 cm (L-band)	FBS: 10X10 FBD: 20X10 PLR: 30X10 ScanSAR: 100	FBS: HH,VV FBD: HH/HV, HH/VH PLR: HH/HV /VH /VV ScanSAR: HH, VV	FBS: 70 FBD: 70 PLR: 30 ScanSAR: 250-350	46	2006-2011	Free	JAXA	High-resolution all-weather soil moisture and ocean surface features observation
RADARSAT-2	5.6 cm (C-band)	Spotlight: ~1.5m	Single: HH, VV, HV, VH	Spotlight: 18x8km	24	2007-	Commercial	CSA	Multi-purpose SAR observation, especially for ice.

		Stripmap: ~3x3-25x25m	Dual: HH/HV, VV/VH	Stripmap: 20-170m					
		ScanSAR: 35x35-100x100m	Quad: HH/HV/VH/VV	ScanSAR: 300x300-500x500km					

Sensors/Satellites	Wavelength	Spatial resolution (m)	No. of bands /Polarization	Swath width (km)	Repeat cycle (days)	Lifetime	Access	Agency	Purpose
TerraSAR-X/TanDEM-X	3.5 cm (X-band)	Spotlight: 0.2x1.0-1.7x3.5m	Single: HH, VV	Spotlight: 3-10	11	2007-	Partial scientific, commercial	DLR	High-resolution all-weather multi-purpose imager for ocean, land and ice. TerraSAR-X and TanDEM-X have coordinated orbits so as to enable computation of accurate Digital Elevation Model
		Stripmap: 3x3m	Dual: HH/VV, HH/HV, VV/VH	Stripmap: 50x30km	11	2010-			
		ScanSAR: 18-40m	Twin: HH/VV, HH/VH, VV/VH	ScanSAR: 150x100-200x200km					
SAR-C (RISAT)	5.6 cm (C-band)	1-50 m	HH or VV or HH/HV or VV/VH	10 to 220 km	7-30	2012-2017; 2022-		ISRO	High-resolution all-weather multi-purpose imager for ocean, land and ice; New EOS 4 with SAR-C to be launched in 2022
COSMO-SkyMed	3.5 cm (X-band)	Spotlight: ≤1m	Single: HH, VV, HV, VH	Spotlight: 10x10km	Satellite: 16 days	2007-	Commercial; limited proposal-based scientific	ASI	High-resolution all-weather multi-purpose imager for ocean, land, and ice
		Stripmap: 3-15m	Dual: HH/HV, HH/VV, VV/VH	Stripmap: 40x40km	Constellation: ~hrs				
		ScanSAR: 30-100m		ScanSAR: 100x100 - 200x200km					

ALOS-2/PALSAR-2	24.6 (L-band)	Spotlight: 1x3m Stripmap: 3-10m ScanSAR: 25-100m	Single: HH, VV, HV, VH Dual: HH/HV, VV/VH Quad: HH/HV/VH/VV	Spotlight: 25x25km Stripmap: 55x70-70x70km ScanSAR: 355x355km	14	2014-	Commercial; limited proposal-based scientific	JAXA	High-resolution all-weather soil moisture and ocean surface features observation
Sensors/Satellites	Wavelength	Spatial resolution (m)	No. of bands /Polarization	Swath width (km)	Repeat cycle (days)	Lifetime	Access	Agency	Purpose
Sentinel-1	5.6 cm (C-band)	Stripmap: 5x5m Interferometric Wide Swath (IW): 5x20m Extra Wide Swath (EW): 20-40m	Single: HH, VV Dual: HH/HV, VV/VH	Stripmap: 375km IW: 250km EW: 400km	12 Constellation: 6 days	2014-	Free	ESA	All-weather ocean and land high resolution multi-purpose observation; SAR-C
SAOCOM	24.6 cm (L-band)	Stripmap: 10x10m TopSAR: 100x100m	Single: HH, VV Dual: HH/HV, VV/VH Quad: HH/HV/VH/VV	Stripmap: >65km TopSAR: 320km	Satellite: 16 days Constellation: 8 days	2019-		CONAE	High-resolution all-weather multi-purpose imager for ocean, land (specifically soil moisture) and ice
SMOS (MIRAS instrument)	(L-band)	50 km	Several polarimetric modes.		3	2010-	Free	ESA	Soil Moisture and Ocean Salinity; Microwave Imaging Radiometer using Aperture Synthesis

SMAP (Soil Moisture Active Passive)	(L-band)	Radiometer: 40 km;	MW radiometer	1000	8	2015-	Free	NASA	Soil moisture in the roots region
		SAR: 30 km (unprocessed, real aperture), 3 km (processed)	Full polarisation						
PAZ SAR	24.6 cm (L-band)	*See TerraSAR/ TanDEM-x	*See TerraSAR/ TanDEM-x	*See TerraSAR/ TanDEM-x	11	2018-	Commercial		
Sensors/Satellites	Wavelength	Spatial resolution (m)	No. of bands /Polarization	Swath width (km)	Repeat cycle (days)	Lifetime	Access	Agency	Purpose
RCM	5.6 cm (C-band)	Very high, high, medium, and low-res modes (3-100m)	Single: HH, VV, VH, HV Dual: HH/HV, VV/VH, HH/VV Compact Quad	20x20-500x500km	Satellite: 12 days	2019-	Commercial	CSA	High-resolution all-weather multi-purpose imager for ocean, land and ice
					Constellation : ~hrs				
NISAR	24.6 cm (L-band)	3-20m (mode dependent)	Single: HH, VV, VH, HV	250	12	2023-	Free	NASA	High-resolution all-weather imagery of ocean and land, especially suited for soil moisture
			Dual: HH/HV, VV/VH, HH/VV						
			Quad						
BIOMASS	69 cm (P-band)	50-60	Quad	50-60	25	2023-	Free	ESA	Mapping of tropical, temperate and boreal forest biomass, including height and disturbance patters
LiDAR sensor									

GEDI	1,064 nm	25	3 lasers are split into 7 beams	7		2018-2023	Free	NASA	Accurate topography and its changes to monitor biomass, ecosystems and ice
Hyperspectral sensor									
Hyperion	0.4-1.0 μm and 0.9-2.5 μm	30	220	7.5	16	2001-2017	Free		Advanced technology for high-resolution land and vegetation observation
GHG emissions retrieval and detection									
Sentinel-5 P (TROPOMI)	270-2,385 nm	7,000	3	2715	Daily	2015	Free	ESA	Tropospheric Monitoring Instrument. Tracked species: BrO, CH4, CO, CO2, H2O, HCHO, NO2, O2, O3, O4, OClO, SO2 and aerosol. Also, solar spectral irradiance.
Sensors/Satellites	Wavelength	Spatial resolution (m)	No. of bands /Polarization	Swath width (km)	Repeat cycle (days)	Lifetime	Access	Agency	Purpose
OCO-2	0.76-2.06 μm	1.29 km (cross-track), 2.25 km (along-track)	3	10	Daily	2019-	Free	NASA	CO2 profile
GOSAT -2 (GOSAT-1)	0.77-14.3 μm	500/1,500	5	1,000/750	6	2018 (2009)	Free	JAXA	Greenhouse gases Observing SATellite-2 "IBUKI-2" equipped with 2 instrument 1. Fourier Transform Spectrometer (TANSO-FTS-2) and 2. Imager (TANSO-CAI-2) to measure CO2, CH4, O3, H2O, CO, NO2

Annex 8. Online tutorials on open source remote sensing software

We provide a sample of online resources for those who are interested to learn more about analysis and processing options of satellite and other geospatial data. We focus exclusively on a few of the most common open source software. A part from online tutorials, Stack Overflow[37] is a 'questions and answers' website for programmers that is very useful to find specific information. Each software also has forums and online community to obtain further guidance. For example, the ESA STEP forum[38] is one place were SNAP users can exchange on analysis. Many researchers and/or institutions publish their code on GitHub, so it is always useful to search for their GitHub page when starting a project. Many tutorials are also available on YouTube, so identifying and subscribing to interesting YouTube channels can be a great source of information.

Table A8.1. Options of online tutorials and training courses on open source remote sensing

Source	Description	GUI/programming language
NASA ARSET	ARSET offers online training modules and webinars that are available on Youtube at introductory, intermediate, and advanced levels on topics relevant to CGIAR activities. The training is also available in Spanish and some in French.	R, python, JavaScript, and others
Copernicus Research and User Support (RUS) service	RUS platform provides online free-access training to promote the uptake of Copernicus data and support capacity-building and research. The tutorials are relevant to CGIAR activities and available in .pdf format as well as via Youtube videos, and use different open source software.	SNAP, R, Python, QGIS
Earth Analytics Courses and Tutorials	Courses and tutorials offered by Earth Lab at University of Colorado, Boulder	R, python, and JavaScript
Regional Agronomy	Edited by Robert Hijmans and Jordan Chamberlin, supported by CIMMYT and the CGIAR Big data platform, this online book provides practical examples related to regional agronomy.	R, python, JavaScript for GEE, and accessing Amazon Web Service
Spatial Thoughts	Founded by Ujaval Gandhi, Spatial Thoughts is a learning platform for geospatial technologies, providing free and online training as well as pay-for instructor-led classes. It provides training on various open source software.	QGIS, Python, Google Earth Engine, and GDAL

[37] https://stackoverflow.com/
[38] https://forum.step.esa.int/

Google Earth Engine tutorials	An online collection of tutorials prepared by several authors for analysis in GEE that touches both fundamentals and applications.	JavaScript
Geemap python package tutorials	Create by Qiusheng Wu, it offers tutorials on GEE and geemap with scripts available on GitHub and Youtube tutorials, and documentation for geemap.	Python
ESA SNAP tutorials	Online tutorials and courses about the use of SNAP toolbox.	SNAP
Spatial Data Science With Applications in R	This book, authored Edzer Pebesma and Roger Bivand, offers a great updated introduction for spatial data manipulation in R.	R
Spatial Data Science with R and "terra"	This online book offered updated information on the new R package 'terra' (that replaces the outdated 'raster' package), which is essential for remote sensing analysis in R	R
Toolkit for Agricultural Geospatial Impact Evaluations	Created by AidData at William & Mary College, this toolkit introduces geospatial impact evaluation for agriculture projects, by supplementing videos, slides, code examples, and other publications.	GEE, Stata
Geo4Dev training	This initiative provides tutorials relevant to impact evaluation using remote sensing.	GEE (python)