Sylvera

How Sylvera uses machine learning (ML)

What is machine learning and why do we use it?

When analyzing carbon projects, Sylvera utilizes machine learning (ML) and multiple types of satellite data to identify specific features of forests and land cover. Examples of these features include:

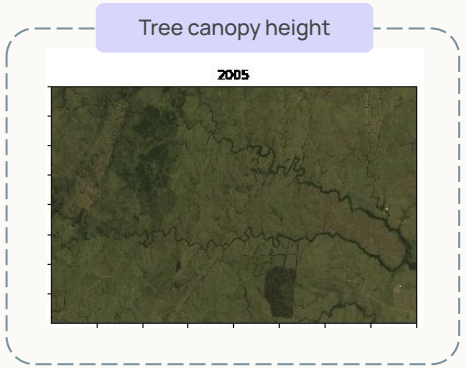
- forest vs. non-forest
- canopy height
- canopy cover
- above ground biomass

Utilizing ML allows us to see what is occurring within project areas (PA) at scale. Rather than manually sampling small areas within a project – which is time consuming and less precise – we can assess whole project areas, located anywhere in the world. In order to have the most accurate output from our ML models, we train proprietary models in specific biomes and geographies, which are used for different carbon project types.

For every PA, we create a shapefile (an outline of the project area). We use the shapefile to extract pixels from satellite imagery that falls in the project boundary (typically 10m-30m resolutions). Sylvera examines every single pixel within a PA both alone and in the context of their adjacencies (which is a powerful feature of deep learning).

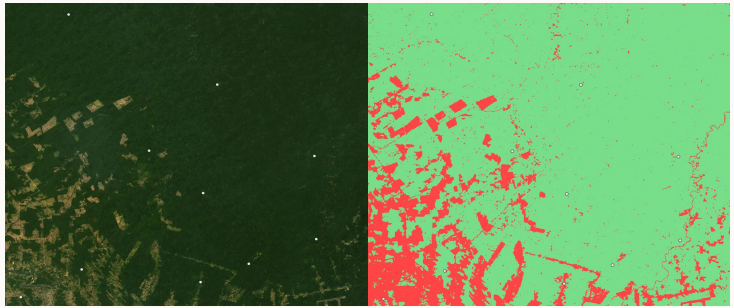
For example, if we are trying to assess forest growth, we will use our ML models to estimate the canopy height of all pixels within the PA. To do this, we train a model to identify forest canopy height by feeding it tens of thousands of labelled data points. This allows the model to "learn" to identify the specific features associated with canopy height.

We then run our models on each carbon project area to estimate – in this example – the canopy height. By running the model over the same area for multiple years, we can see changes over time in the forest area.



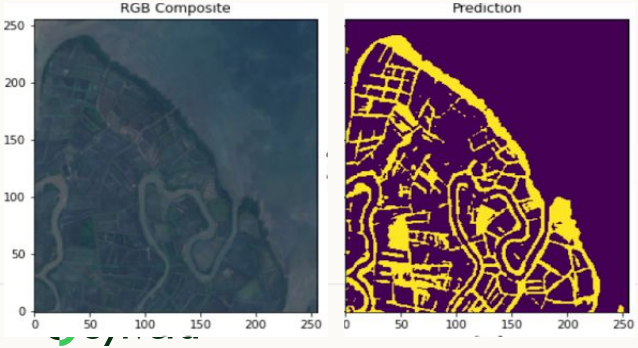
Our Geographic Information System (GIS) and ML capabilities lead the market

Forest/non forest

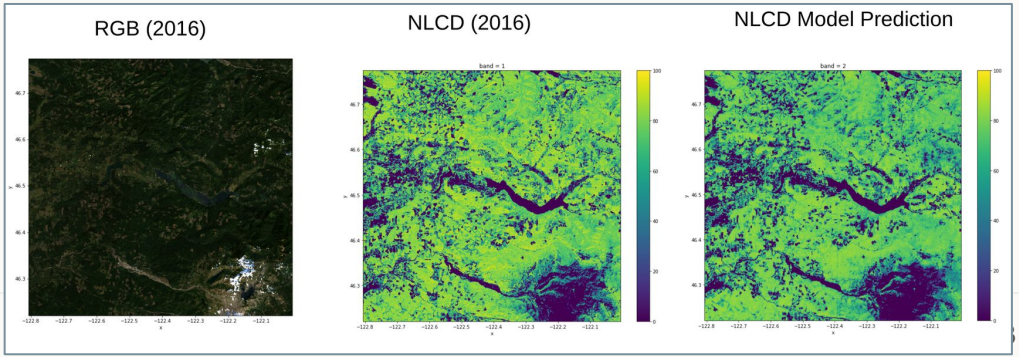


Tree canopy height

Mangroves



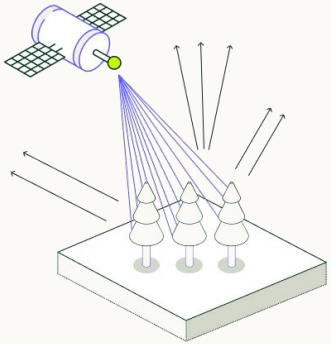
Tree canopy cover



We combine multiple data sources

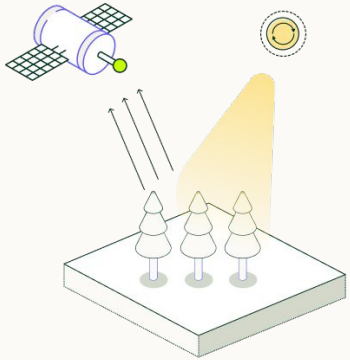
We utilize multiple types of data to train and run our models. Each type provides different data, enabling us to detect specific features.

Synthetic Aperture Radar (SAR) Satellites



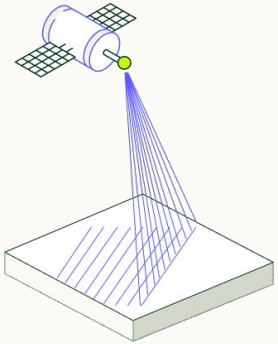
E.g. ALOS PALSAR, Sentinel-1

Optical Satellites



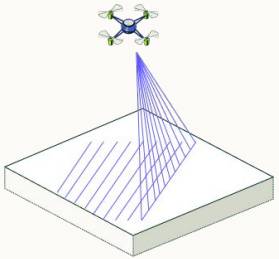
E.g. Landsat-7, Landsat-8, Sentinel-2

LiDAR Satellites



E.g.GEDI

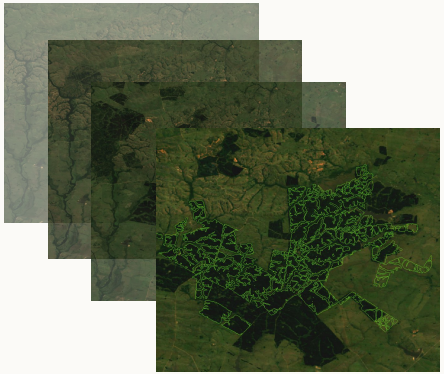
Multi-Scale LiDAR



E.g. proprietary terrestrial and UAV lidar data

Our proprietary models are run on project areas to estimate forest area, canopy height, etc.

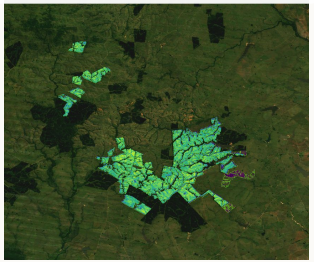
DATA COLLECTION

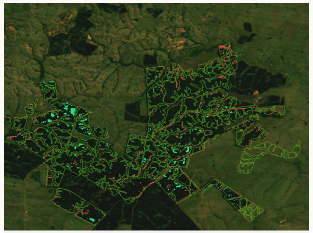


Satellite data collected over years relevant to project area

PROCESSING

Time-stamped predictions made for project area e.g. canopy height

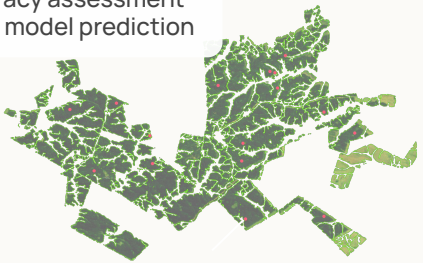


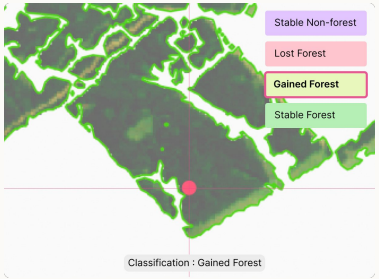


Combine predictions over time to identify areas of forest growth or loss

FINALIZING DATA

Accuracy assessment run on model prediction





Validated predictions, e.g. canopy height over time are fed into our ratings.

Example 1: Tree Canopy Height in ARR Project

ARR (Afforestation, Reforestation, Revegetation) projects are nature-based solutions that fall under the "removals" category. Afforestation projects aim to convert degraded and barren land through tree planting.

To evaluate the performance of ARR projects, it is essential to identify areas of new forest growth, as well as forest loss. We use canopy height as a proxy to identify these areas.

The images on the following page illustrate how we utilize our proprietary machine learning models and satellite data to identify areas of forest growth by estimating canopy height within the ARR project areas over time.

In this example, you can see the canopy height increases in different areas of the project. This indicates areas of new forest planting and growth at different points in time.

How does this inform the Sylvera credit rating?

We compare these results to the project reports to identify if they align or if there are discrepancies.

- If we identify the same area of forest growth as reported, and no unreported losses, then the carbon score will be 100%
- If we identify a smaller area of forest growth than reported, or any forest loss that is unreported, then the carbon score will be below 100%.

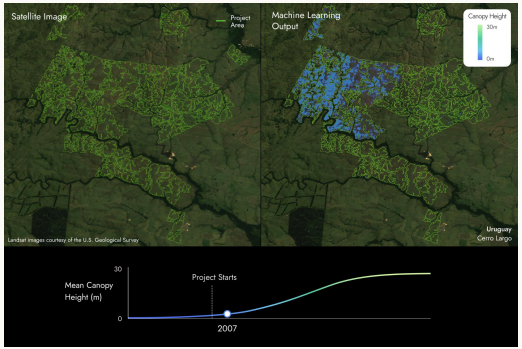
The results of our modelling approach are put through a quality control process to ensure that they are representative of what is actually happening on the ground within the project area, and to ensure a high accuracy in the Carbon Score.

We also use our ML results in the Additionality component of our ratings framework to test for Over-Crediting Risk.

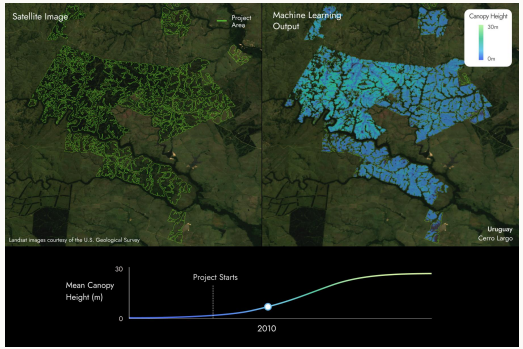
- If a project developer has cleared *any* primary forest prior to the project start, then the area of primary forest cleared is considered ineligible and represents an over-crediting risk. To help us quantify this risk, we use our ML results to track the extent of any historic primary forest loss prior to project start.

Example 1: Tree Canopy Height in ARR Project

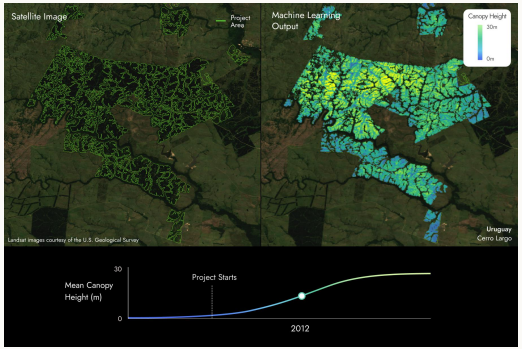
2007



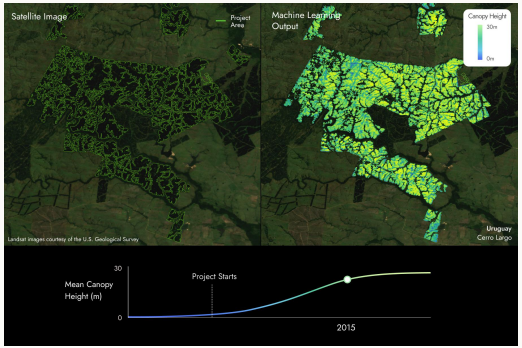
2010



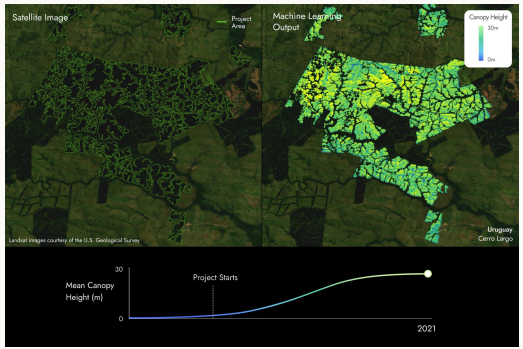
2012



2015

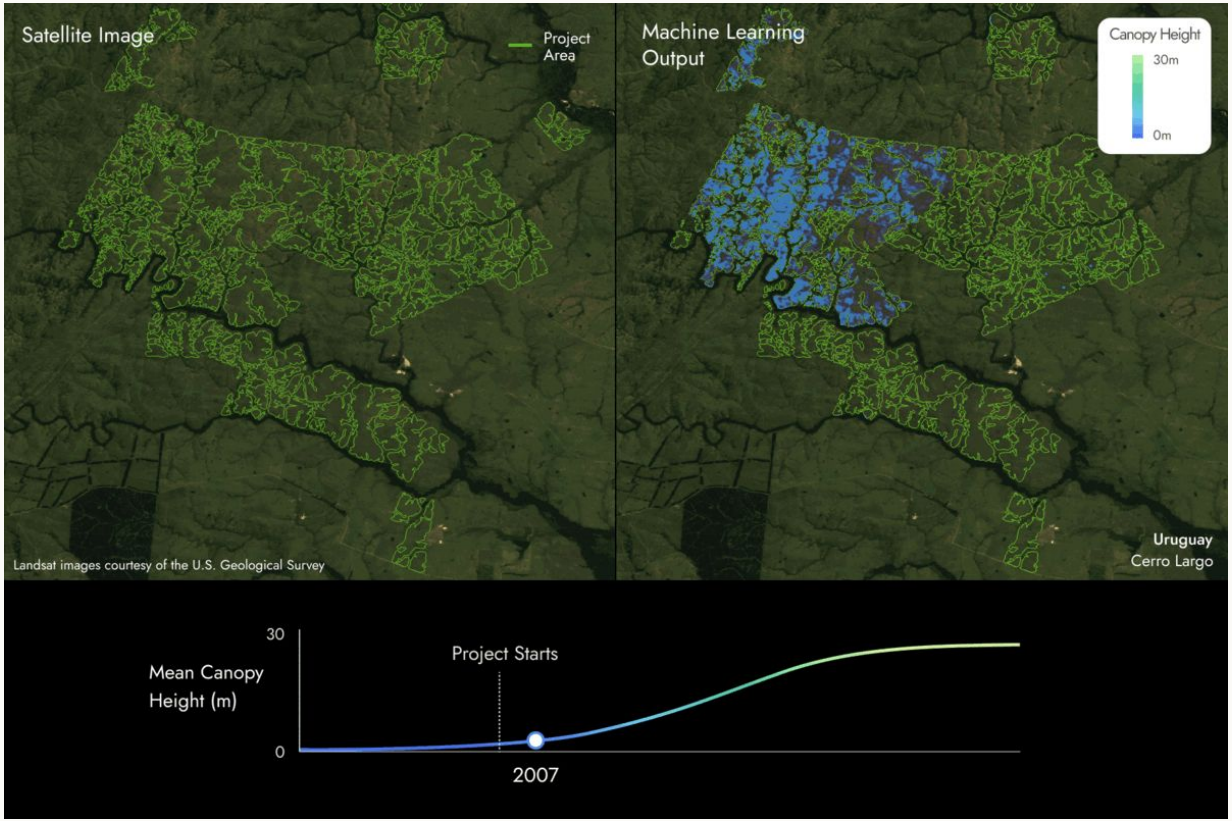


2021



Canopy height (yellow and green) increases in different areas of the project over time. This indicates areas of new forest planting and growth.

Example 1: Tree Canopy Height in ARR Project



Example 2: REDD+ Project

REDD+ (Reduce Emissions from Deforestation and Forest Degradation) projects are a common type of nature-based solution that fall within the “avoidance” category. They aim to preserve existing forests by attaching financial value to the carbon stored in forests, thereby incentivizing forest conservation.

To evaluate the performance of REDD+ projects, it is necessary to identify forest loss over a period of time.

The images on the following page show how we utilize our proprietary machine learning models and satellite data to identify areas of forest loss by detecting whether an area changes from forest to non-forest within REDD+ project areas across time.

The results of our modelling approach are put through a quality control process to ensure that they are representative of what is actually happening on the ground within the project area, and to ensure sufficient accuracy levels in carbon score.

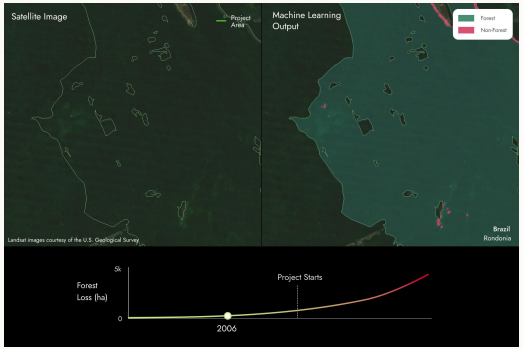
In this case, you can see that there are some areas changing from forest to non forest, indicating that forest loss has occurred within the project area over time.

We then compare these results to the reported forest loss in project documentation to identify whether they align or if there are discrepancies.

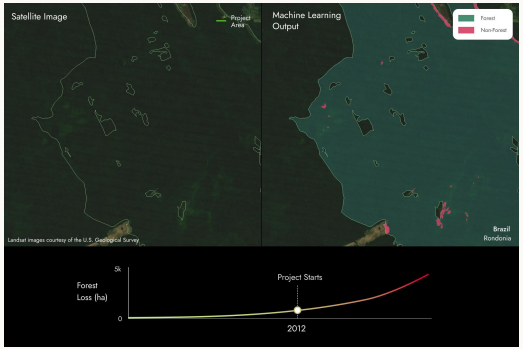
- If we find that there is more forest loss than reported, then the carbon score will be below 100%.
- The more forest loss that is not reported, the lower the carbon score.

Example 2: REDD+ Project

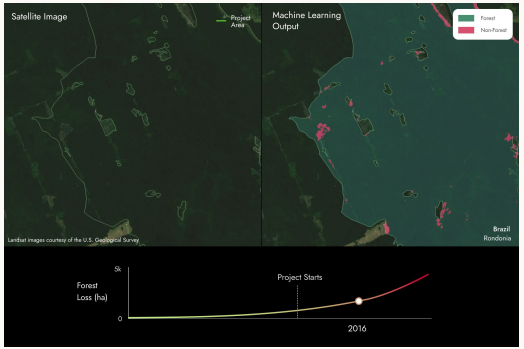
2006



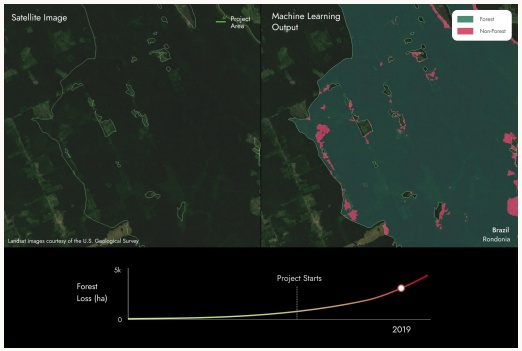
2012



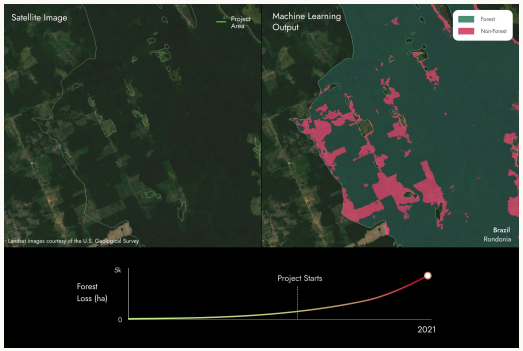
2016



2019

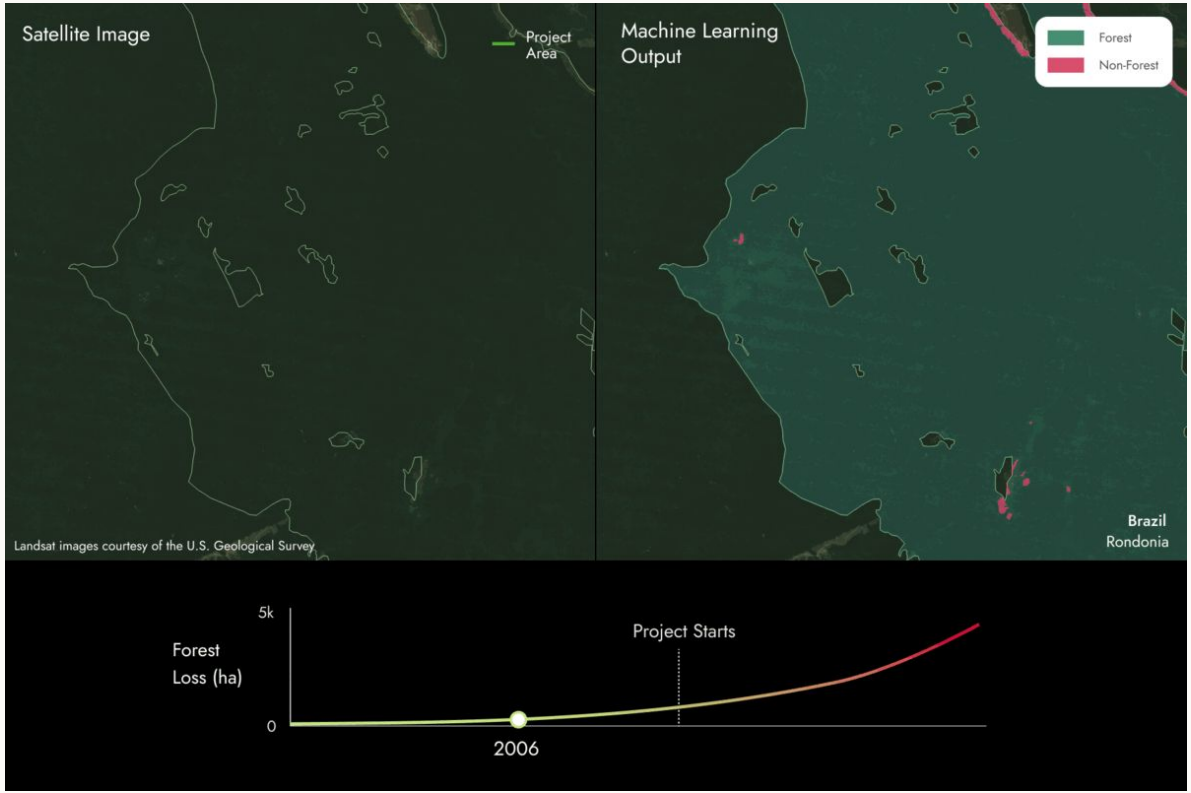


2021



The increase in non forest areas (red) within the PA indicate that forest loss has occurred.

Example 2: REDD+ Project



How does this make us different from competitors?

Sylvera's machine learning expertise and capabilities gives us the ability to accurately gain information on carbon projects that others cannot. We also invest into future research streams to ensure we are always operating at the cutting edge of the industry.

In addition to classical machine learning, which assesses each data point (i.e., pixel) in isolation, Sylvera also applies deep learning algorithms, which are able to better understand image context. This enables greater accuracy in estimation and analysis, and for generalized prediction across different geographies and time periods.

Sylvera does not rely on optical data alone, but makes use of the full range of EO data available: optical, SAR/Radar, and LiDAR. Soon we'll be incorporating hyperspectral data into our methodology.

Ground Truthing Campaign

Sylvera is LiDAR scanning forests in an effort to build the world's largest dataset of carbon stored in trees and above-ground biomass.

The quantity and quality of this LiDAR reference data is unparalleled. By collecting vast quantities of this data around the world in different biomes, it means that Sylvera can estimate both biomass and carbon stocks for forests at an unprecedented accuracy using satellite data.

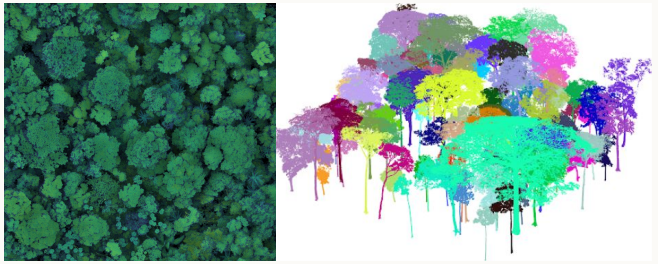


Figure 1. Point clouds created from Sylvera LiDAR scans.

Sylvera is the only ratings platform in the world that has collected enough three dimensional forest data using LiDAR for it to be used as reference data. Today, we use this data to train and calibrate the EO-based machine learning models to improve their accuracy. This is one of our key research and development streams and will be used in our ratings in the future.

Example 3: Tree Canopy Cover and Activities in IFM Project

Improved Forest Management (IFM) is an agriculture, forestry and land-use project category. These forest management activities result in increased carbon stocks within forests and/or reduce greenhouse gas emissions from forestry activities when compared to business-as-usual forestry practices.

To evaluate the performance of IFM projects, we identify activities that have occurred that result in a change in carbon stock. We use canopy cover as a proxy for these activities and a combination of deep learning and regression models to identify the size of the area in which these activities have taken place.

The image on the following page illustrates how we utilize our proprietary machine learning models and satellite data to identify changes in forest. In this example, you can see a map indicating the relative change in forest cover in different areas of a project - these changes in forest cover will have a relative impact on the carbon stock.

How does this inform the Sylvera credit rating?

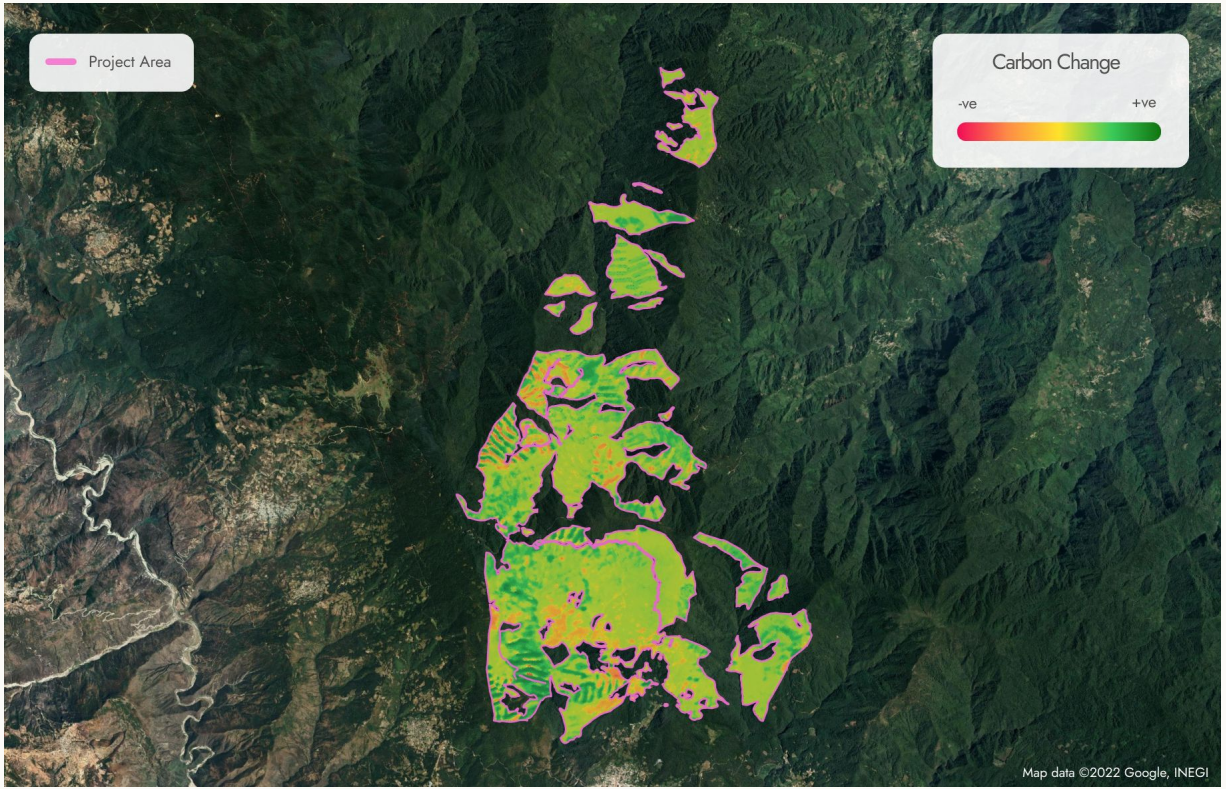
The outputs of our models are put through a robust human-driven quality control process to ensure that they are representative of what is actually happening on the ground within the project area and to ensure a high accuracy in the Carbon Score.

We compare the changes in forest we've identified to the project reports to determine if they align or if there are discrepancies between what we observe and what is claimed by the project developers (e.g. amount of forest harvested).

- In the example you can see that there are areas of both negative and positive impact on carbon from the project start to the latest monitoring report.
- If the project is reporting less carbon loss than we detect, such as from harvesting, then this will lead to a lower carbon score.

The same method is used from 1993 to 2021 for a buffer area around the project and within the project area. This feeds into the additionality component of our rating and allows us to compare forest behavior in the project area and in the area surrounding it.

Example 3: Tree Canopy Cover and activities in IFM Project



Map indicating areas with a positive or negative impact on carbon stock for a project.

Sylvera is the leading carbon credit ratings platform. We help corporate sustainability leaders, traders and asset managers confidently evaluate and invest in the best carbon credits. By creating the first carbon intelligence platform, Sylvera is raising the bar on project accounting and analysis, and introducing a much needed source of truth for carbon markets. We are backed by renowned investors like Index Ventures, Insight Partners, LocalGlobe and Salesforce Ventures.

To learn more about Sylvera, contact us.

Sylvera