Approach and methodology for Terra-i+

Dataset and metric description

Prepared by the Alliance of Bioversity International and International Center for Tropical Agriculture

TABLE OF CONTENTS

I INTRODUCTION

This methodology document aims to provide transparency and accessibility for users of the Terra-i+ webtool. Section I offers background information and outlines the webtool's objectives. Section II details the methodology for calculating the metrics displayed in the webtool, while Section III explains the methodology and data sources used to create the geospatial dataset supporting these metrics.

THE CHALLENGE: Growing demands for geospatial data, driven by voluntary sustainability standards and the EU Deforestation-Free Regulation (EUDR) traceability requirement, pose significant challenges. Additionally, increased commitments to carbon monitoring at the farm level heighten the need for accurate carbon stock data. The scarcity of reliable coffee and forest location data complicates efforts to mitigate the risk of coffee production being linked to deforestation. Moreover, the absence of an objective source on forest-to-coffee conversion timing, coupled with the costliness of collecting and updating carbon footprint data over time, exacerbates the challenge.

THE SOLUTION: Terra-i+ addresses these challenges by utilizing high-resolution imagery from Sentinel 1&2 and AI technology to generate accurate land cover maps for coffee and forest layers, along with biweekly deforestation alerts and shade/visible soil coverage maps. Furthermore, a future climate suitability map was constructed using historical climate data and a machine learning classification method. Terra-i+ extracts key information from these data layers, creating a user-friendly online platform for agribusinesses. This platform empowers agribusinesses to manage deforestation risks in their supply chain, meet carbon commitments, and actively contribute to global deforestation and carbon footprint reduction.

II METRIC CALCULATION

Terra-i+ offers three comprehensive dashboards, each designed to provide in-depth metrics for the evaluation of environmental risks at both jurisdictional and farm levels. These metrics encompass critical aspects such as commodity-driven deforestation, adverse effects on water bodies, risks associated with climate change, and opportunities for carbon sequestration through agroforestry practices. The development of each metric is grounded in robust scientific methodologies and innovative datasets derived from remote sensing and AI technologies. An integral aspect of metric creation involved an iterative process of engagement and co-creation with target users within the coffee industry. Placing users at the core of the design process, each metric was tailored to their specific needs, following a collaborative approach illustrated in Figure 1. This user-centric approach ensures the relevance and effectiveness of the metrics in addressing the unique challenges and requirements of the coffee industry stakeholders.

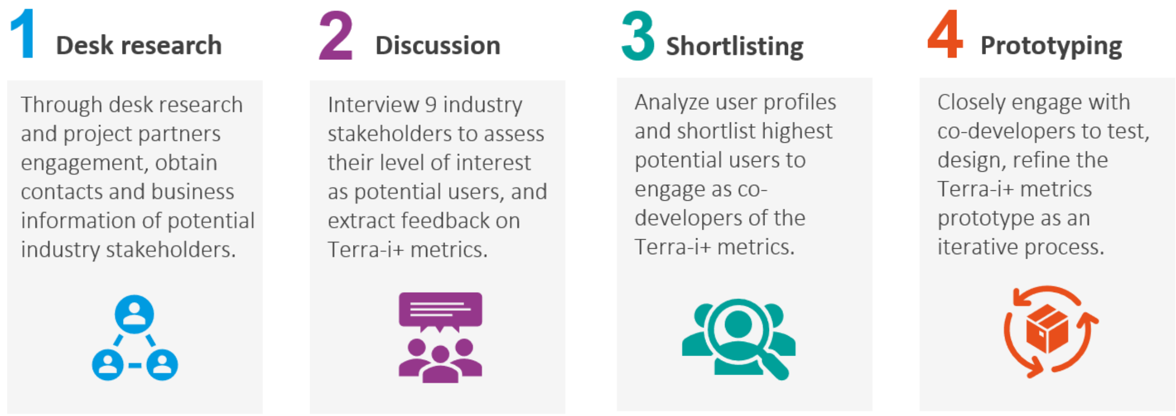


Figure 1 User-centric design approach to develop the metrics.

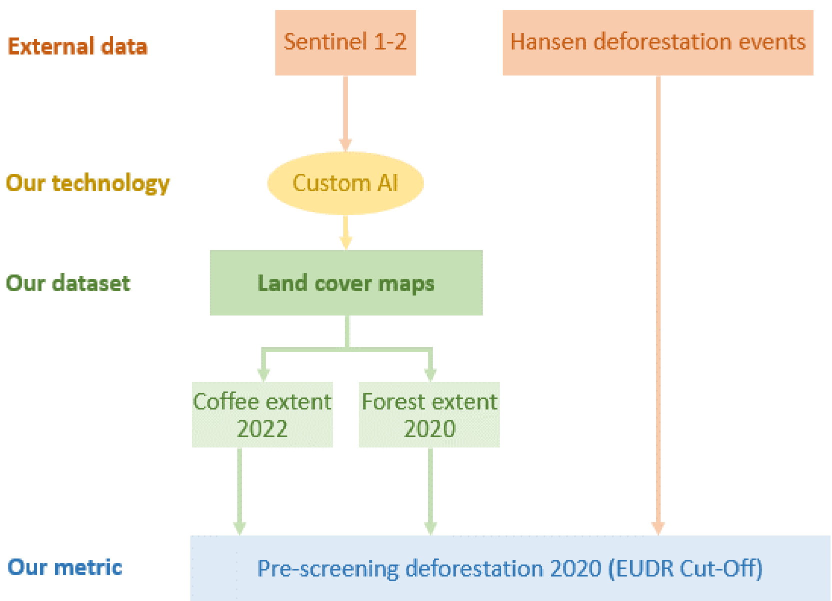
The following sections describe the process through which each metric is calculated.

1 Dashboard: Jurisdiction level evaluation

The "Dashboard: Jurisdiction level evaluation" helps users benchmark prospective areas against each other. It doesn't require any input data from the user. The metrics within this dashboard are aggregated at the jurisdiction level selected by the user.

1.1 Metric: Pre-screening deforestation 2020 (EUDR Cut-Off)

Pre-screening deforestation 2020 (EUDR Cut-Off) metric is used to benchmark deforestation risks of prospective sourcing areas, to evaluate risk of non-compliance with EU Deforestation Regulation.



First, all Hansen deforestation events (see section III.5. External datasets) are filtered to restrict the selection to deforestation **after the EUDR cut-off (31 Dec 2020)**. For each selected jurisdiction unit, a likelihood level is calculated based on the following criteria:

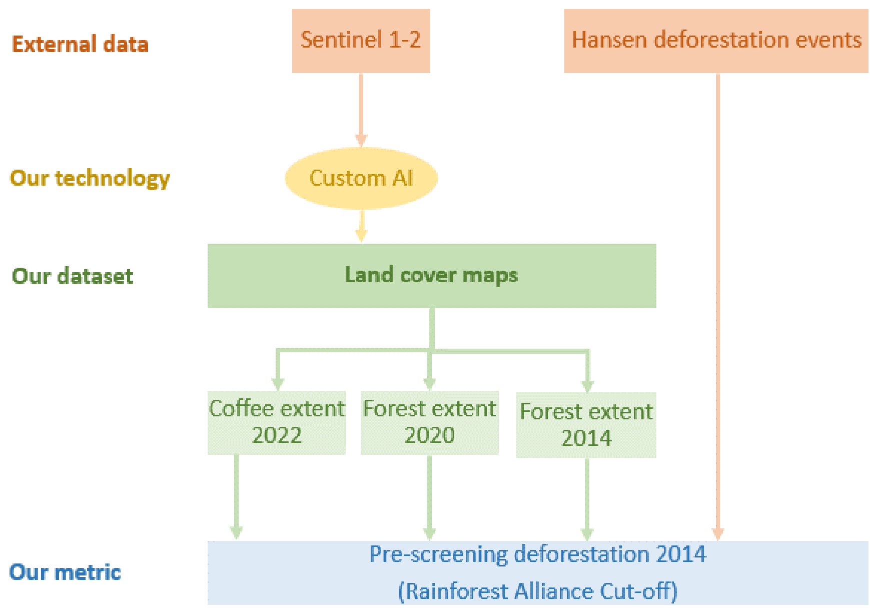
- **Very high likelihood coffee driven deforestation**: A zonal statistic in hectare of the intersection of deforestation events (Hansen dataset), forest in 2020 (from Forest map 2020 – see section III.1. Land cover), current coffee area from the most up to date version of land cover (see section III.1. Land cover).
- **High likelihood coffee driven deforestation:** A zonal statistic in hectare from the result of filtering all the deforestation events to keep only the ones that happened within the last 2 years, are located within forest extent of 2020 and are located no more than 200 meters from a coffee area but not within a coffee area (from the most up to date version of land cover). This likelihood level thus captures the risk of recent deforestation linked to newly planted coffee undetectable by remote sensing, close to the existing coffee land cover.
- **Moderate likelihood coffee driven deforestation**: A zonal statistic in hectare from the result of filtering all the deforestation events to keep only the ones that happened within the last 2 years, are located within forest extent of 2020 and are located 200-500 meters from a coffee area but not within a coffee area (from the most up to date version of land cover). This likelihood level thus captures the risk of recent deforestation linked to newly planted coffee undetectable by remote sensing, close to the existing coffee land cover.
- **Low likelihood coffee driven deforestation**: A zonal statistic in hectare of the deforestation events located within forest extent of 2020 and are located more than 500 meters from a coffee area (from the most up to date version of land cover) or happened more than 2 years ago (in other word, all the deforestation that happened within forest in 2020 and are not coffee driven deforestation and not potentially coffee driven deforestation).

The results are displayed in two ways:

- **Absolute value:** hectare of deforestation
- **Relative value:** hectare of deforestation as percentage of 2020 forest area

1.2 Metric: Pre-screening deforestation 2014 (Rainforest Alliance Cut-off)

Pre-screening deforestation 2014 (Rainforest Alliance Cut-off) metric optimizes the certification selection process by benchmarking prospective sourcing areas against each other, to evaluate deforestation risk since 2014 (Rainforest Alliance Cut-Off).



First, all Hansen deforestation events (see section III.5. External datasets) are filtered to restrict the selection to deforestation **since Jan 1, 2014**. For each selected jurisdiction unit, a likelihood level is calculated based on the following criteria:

- **Very high likelihood coffee driven deforestation**: a zonal statistic in hectare of the intersection of deforestation events (Hansen dataset), forest in 2014 (from Forest map 2014 – see section III.1.6. 2014 Land cover map), current coffee area from the most up to date version of land cover map (see section III.1. Land cover).

- **High likelihood coffee driven deforestation:** A zonal statistic in hectare from the result of filtering all the deforestation events to keep only the ones that happened within the last 2 years, are located within forest extent of 2020, and are located no more than 200 meters from a coffee area but not within a coffee area (from the most up to date version of land cover). This likelihood level thus captures the risk of recent deforestation linked to newly planted coffee undetectable by remote sensing, close to the existing coffee land cover.

- **Moderate likelihood coffee driven deforestation:** A zonal statistic in hectare from the result of filtering all the deforestation events to keep only the ones that happened within the last 2 years, are located within forest extent of 2020, and are located 200-500 meters from a coffee area but not within a coffee area (from the most up to date version of land cover). This likelihood level thus captures the risk of recent deforestation linked to newly planted coffee undetectable by remote sensing, close to the existing coffee land cover.

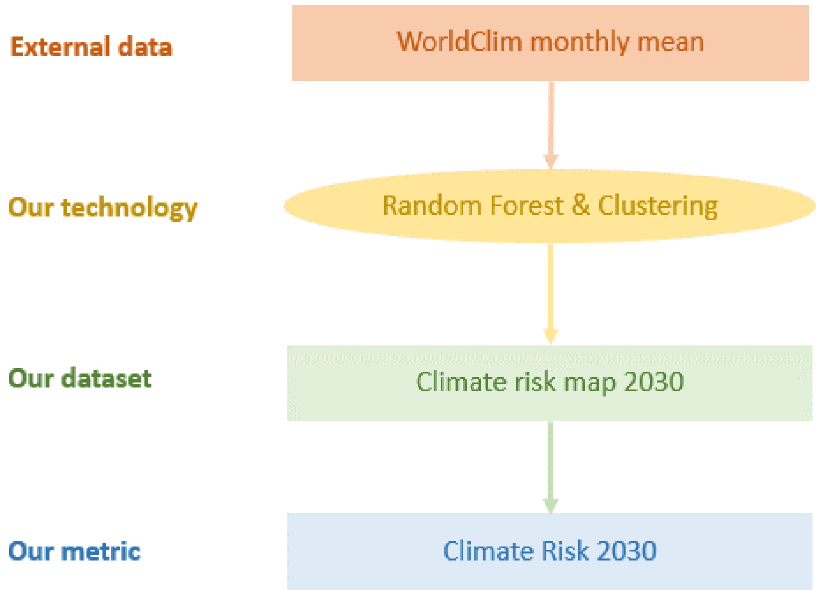
- **Low likelihood coffee driven deforestation**: A zonal statistic in hectare of the deforestation event located within forest extent of 2014, and are located more than 500 meters from a coffee area (from the most up to date version of land cover) or happened more than 2 years ago (in other word, all the deforestation that happened within forest in 2014 and are not coffee driven deforestation and not potentially coffee driven deforestation).

The results are displayed in two ways:

- **Absolute value:** hectare of deforestation
- **Relative value:** hectare of deforestation as percentage of 2014 forest area

1.3 Metric: Climate risk 2030

Climate risk 2030 metric (jurisdiction level) identifies jurisdictions in need of adaptation intervention by assessing future climate suitability for coffee growing.



For each selected jurisdiction unit, a zonal statistic of the hectare within the coffee land cover in hectare (see section III.1) that falls into different climate suitability levels was computed. The suitability levels are derived from the climate suitability map (see section III.4 on how the climate suitability map was computed). The climate suitability levels are:

- Remain/become suitable: Areas where the climate is most likely to remain or become suitable for Robusta coffee production.
- High adaptation needs: Areas where climate change will likely cause substantial stress in traditional Robusta production systems. Adaptation such as adapted varieties, diversification, and financial mechanisms will be necessary to reduce risk of becoming unsuitable.
- Highest adaptation needs: Areas where climate is more likely to make coffee production less feasible. Adaptation will require redesigning the production system or switching to new crops.
- Remain/become unsuitable: Areas where climate is most likely to remain or become unsuitable for Robusta coffee production.

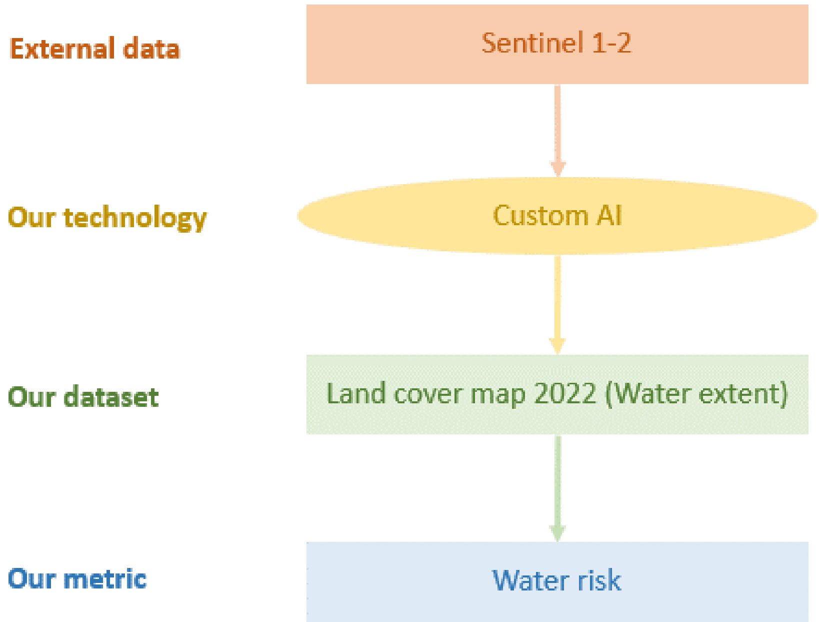
The results are displayed in two ways:

- **Absolute value:** hectare of coffee in each climate suitability level.

- **Relative value:** hectare of coffee in each climate suitability level, as percentage of total coffee area.

1.4 Metric: Water risk

Water risk metric (jurisdiction level) evaluates which jurisdictions have more coffee area close to surface water bodies, presenting a potential risk to water quality.



For each selected jurisdiction unit, a zonal statistic in hectare of the intersection of a 50m buffer around all the surface water bodies and the current coffee area from the most up to date version of land cover map (see section III.1.) is computed.

The results are displayed in two ways:

- **Absolute value:** hectare of coffee within (≤) 50m of water bodies.
- **Relative value:** hectare of coffee within (≤) 50m of water bodies, as percentage of total coffee area.

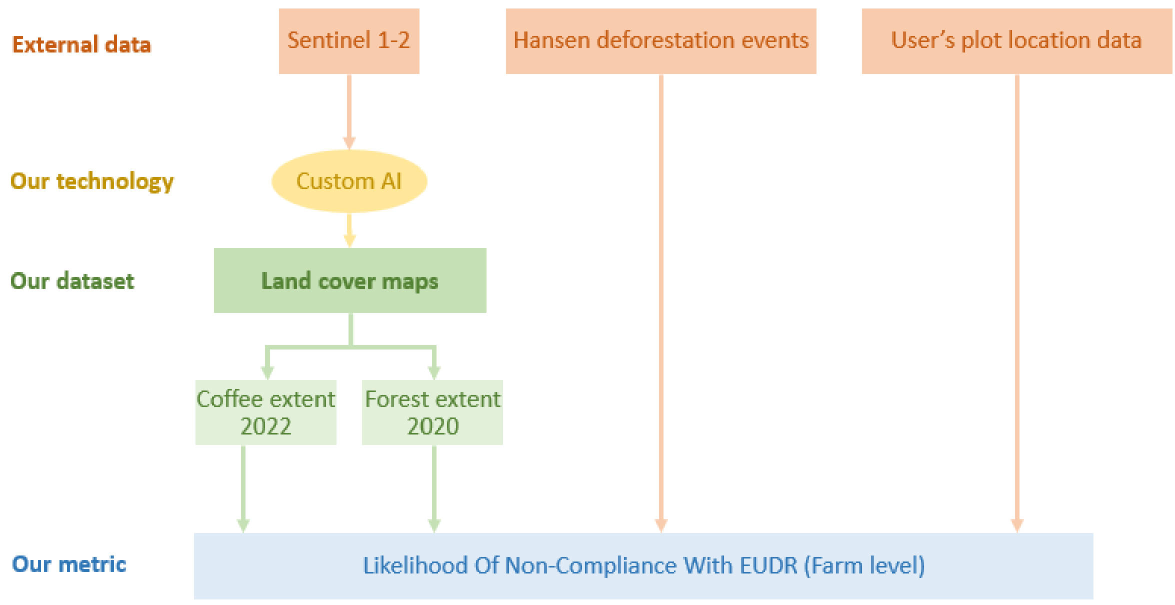
2 Dashboard: Farm level compliance

The "Dashboard 2: Farm level compliance" helps users identify which farm(s) in their supply chain are at risk. Therefore, this dashboard uses plot location data from the user as input data. The metrics within this dashboard will be calculated at plot level and then aggregated to farm (farmer) level.

The plot units provided by the users can either be polygons or single GPS locations. If the plot unit is represented as a single GPS location, a buffer (circle) corresponding to the given farm size will be computed and used as approximation of the plot location.

2.1 Metric: Likelihood Of Non-Compliance With EUDR

Likelihood Of Non-Compliance with EUDR metric pinpoints farm level deforestation risks to evaluate likelihood of non-compliance with the EU Deforestation Regulation.



For each plot unit, a risk level is calculated. First, all Hansen deforestation areas (see section III.5. External datasets) are filtered to restrict the selection to deforestation **after the EUDR cut-off (31 Dec 2020)**. The plot level likelihood of **non-compliance with EUDR** is then computed as below:

- **Very high likelihood:** Plots with deforestation inside the plot boundary
- **High likelihood:** Plots with deforestation that lies within 0-200m from plot boundary, and the deforestation is either (1) within coffee land cover (on the most recent land cover map) OR (2) not within coffee land cover (on the most recent land cover map) but occurs in the last 2 years.
- **Moderate likelihood:** Plots with deforestation that lies within 200-500m from plot boundary, and the deforestation is either (1) within coffee land cover (on the most recent land cover map) OR (2) not within coffee land cover (on the most recent land cover map) but occurs in the last 2 years.
- **Low likelihood:** Plots that do not belong to the above categories.

The farm's non-compliance likelihood level is equal to the highest likelihood level among its plots.

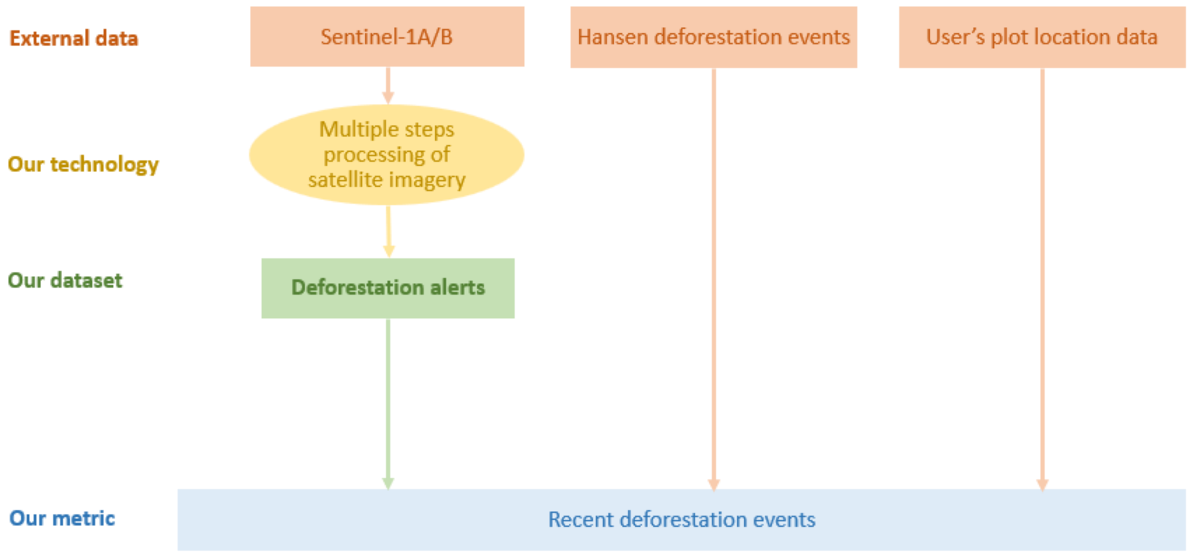
Details on the methodology to map the forest land cover and coffee land cover can be found in section III.1. Land cover.

The results of this metric are displayed in two ways:

- **Farm area:** total area of farms under each non-compliance likelihood.
- **Farm count:** number of farms under each non-compliance likelihood.

2.2 Metric: Recent Deforestation Events

Recent Deforestation Events metric pinpoints farm level risk of being linked to deforestation within the last 1 year.



This metric assesses the risk of being linked to deforestation by identifying farms within 500m of last 12 months deforestation events (see section III.3. Deforestation monitoring system) and putting a higher priority on farms that has deforestation events nearer to the farm, and deforestation in protected areas.

First, deforestation alerts are filtered to restrict the selection to the last 12 months. For each plot unit, a risk level is calculated based on the following criteria:

- ***High priority plots – within Protected area:*** *The deforestation events within 500m of the plot boundary lie within protected areas.*
- ***High priority plots:*** *The deforestation events within 500m of the plot boundary are NOT within protected area, and satisfy one of these requirements: (1) the events lie within 0m-100m of plot boundary; OR (2) there are more than 4 events within 500m of the plot.*
- ***Medium priority plots:*** *The deforestation events within 500m of the plot boundary are NOT within protected area, and satisfy one of these requirements: (1) the events lie within 100m-200m of plot boundary; OR (2) there are 2-4 events within 500m of the plot.*
- ***Low priority plots:*** *Plots that have deforestation events within 500m of the plot boundary, but don't belong to the above categories.*
- ***No risk plots:*** *Plots that have no deforestation events within 500m of the plot boundary.*

	No risk	**Low risk**	**Medium risk**	**High risk**	**High risk – Within Protected Area**
Proximity from alert(s) to plot	≥500m	200m-500m	100-200m	<100m	<500m

Number of alert(s) within 500m the plot	0	<2	2-4	>4	≥ 1
Whether any alert(s) are inside protected area	Not applicable	No	No	No	Yes

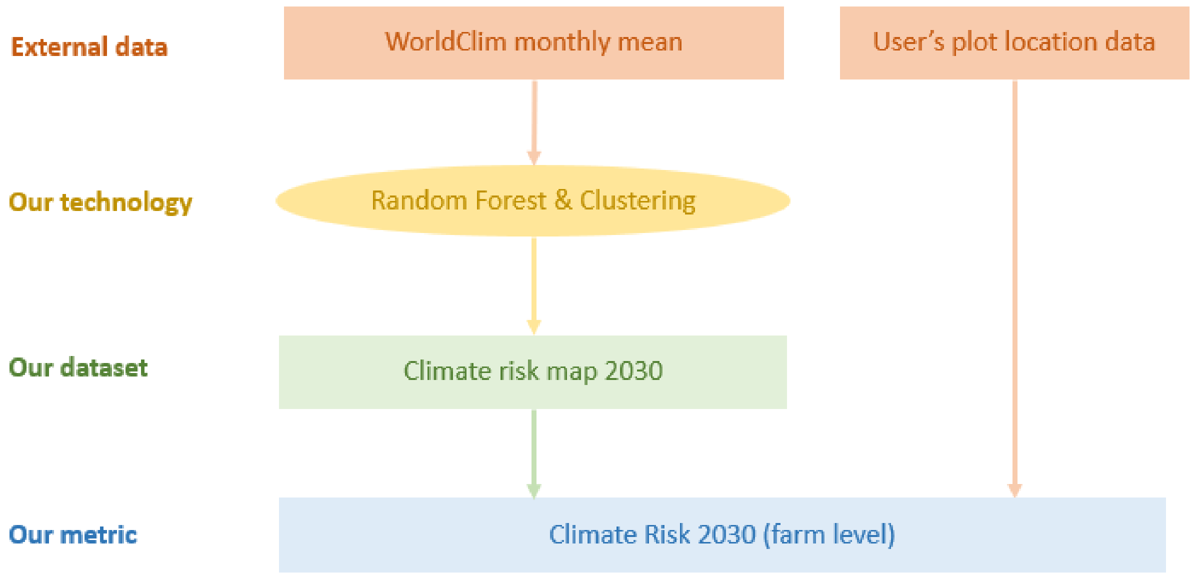
The farm's priority level is equal to the highest priority level among its plots.

The results of this metric are displayed in two ways:

- **Farm area:** total area of farms under each risk level.
- **Farm count:** number of farms under each risk level.

2.3 Metric: Climate Risk 2030

Climate Risk 2030 metric (farm level) pinpoint farms in need of adaptation intervention by assessing future climate suitability for coffee growing.



For each farm, the zonal statistic of the area within the coffee land cover in hectare (see section III.1) that falls into different climate suitability levels is computed. The suitability levels are derived from the climate suitability map (see section III.4 on how the climate suitability map was computed). The climate suitability levels are:

- Remain/become suitable: Areas where the climate is most likely to remain or become suitable for Robusta coffee production.
- High adaptation needs: Areas where climate change will likely cause substantial stress in traditional Robusta production systems. Adaptation such as adapted varieties, diversification, and financial mechanisms will be necessary to reduce risk of becoming unsuitable.

- Highest adaptation needs: Areas where climate is more likely to make coffee production less feasible. Adaptation will require redesigning the production system or switching to new crops.
- Remain/become unsuitable: Areas where climate is most likely to remain or become unsuitable for Robusta coffee production.

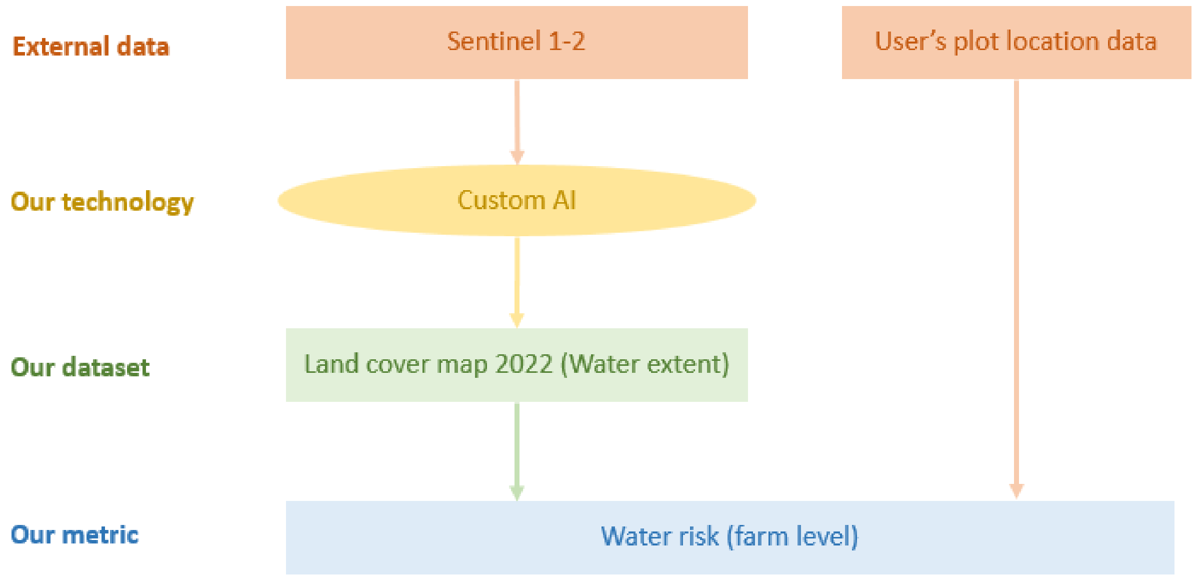
The climate risk level assigned to the farm corresponds to the risk level covering the largest farm area.

The results of this metric are displayed in two ways:

- **Farm area:** total area of farms under each climate suitability level.
- **Farm count:** number of farms under each climate suitability level.

2.4 Metric: Water Risk

Water risk metric (farm level) pinpoints farms that are very close to surface water bodies, presenting a potential risk to water quality.



For each plot unit, a binary risk level is assigned based on whether the plot unit is located within 50m from a surface water body. The farm is considered at risk if any of its plots are located <50m from a water body. Distances are calculated with a simple vector operation.

The map of surface water bodies is derived from the most up to date version of land cover map (see section III.1.)

The results of this metric are displayed in two ways:

- **Farm area:** total area of farms near water (<50m from water bodies), and far from water (≥50m from water bodies).
- **Farm count:** number of farms near water (<50m from water bodies), and far from water (≥50m from water bodies).

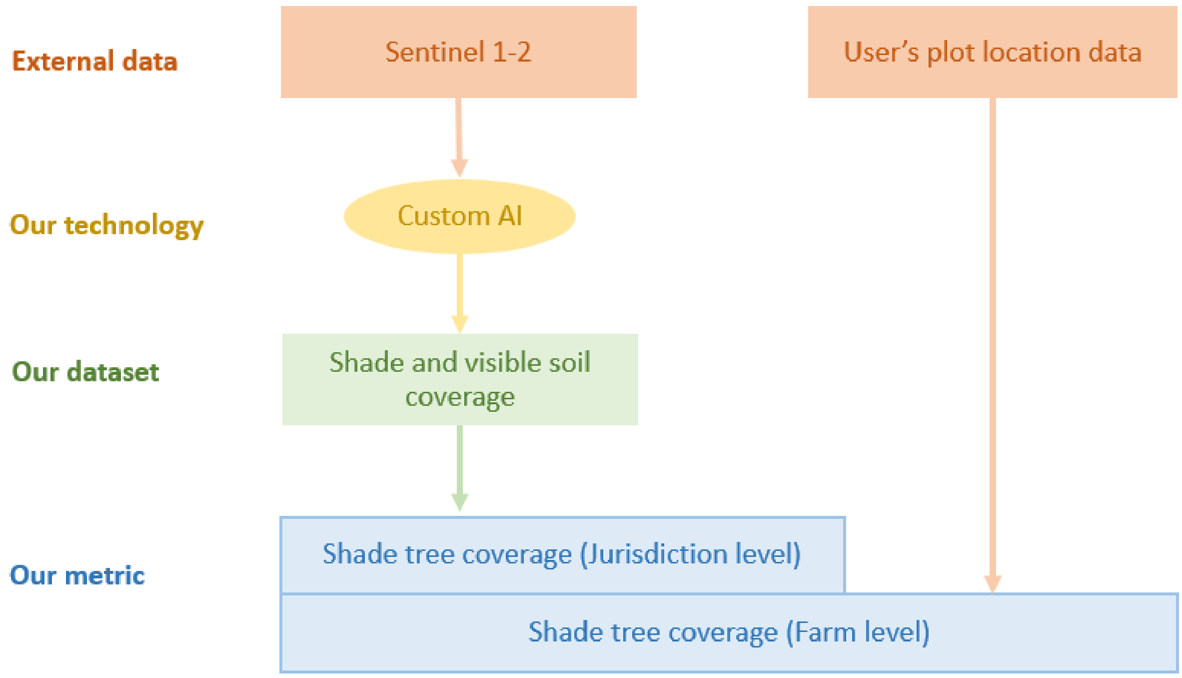
3 Dashboard: Carbon project opportunity

This dashboard can be used either to assess at (1) jurisdiction level, or (2) at farm level within the users' supply chain.

It provides an assessment of the current shade tree coverage on coffee farm and potential for carbon gain in agroforestry intervention, when the current coffee shade system is fully converted into heavy shaded coffee (>15% coverage) within selected areas.

3.1 Metric: Shade tree coverage

Shade tree coverage metric benchmarks jurisdictions and farms by area of coffee under different shade levels (no shade, light shade, moderate shade, and heavy shade).



This metric can be used at either (1) jurisdiction level, or (2) farm level.

Shade levels are derived from shade and visible soil map (see section III.2.), and are categorized as follows, with % means shade canopy cover percentage:

- 0 – 5%: No shade
- 5% - 10%: Light shade
- 10% - 15%: Moderate shade
- >15%: Heavy shade

For jurisdiction level, the results are computed and displayed in two ways:

- **Absolute value:** hectare of coffee under each shade level.
- **Relative value:** hectare of coffee under each shade level, as percentage of total coffee area.

For farm level, the results are computed and displayed in two ways:

- **By coffee area**: The number of hectares for each shade level, within the selected farms.

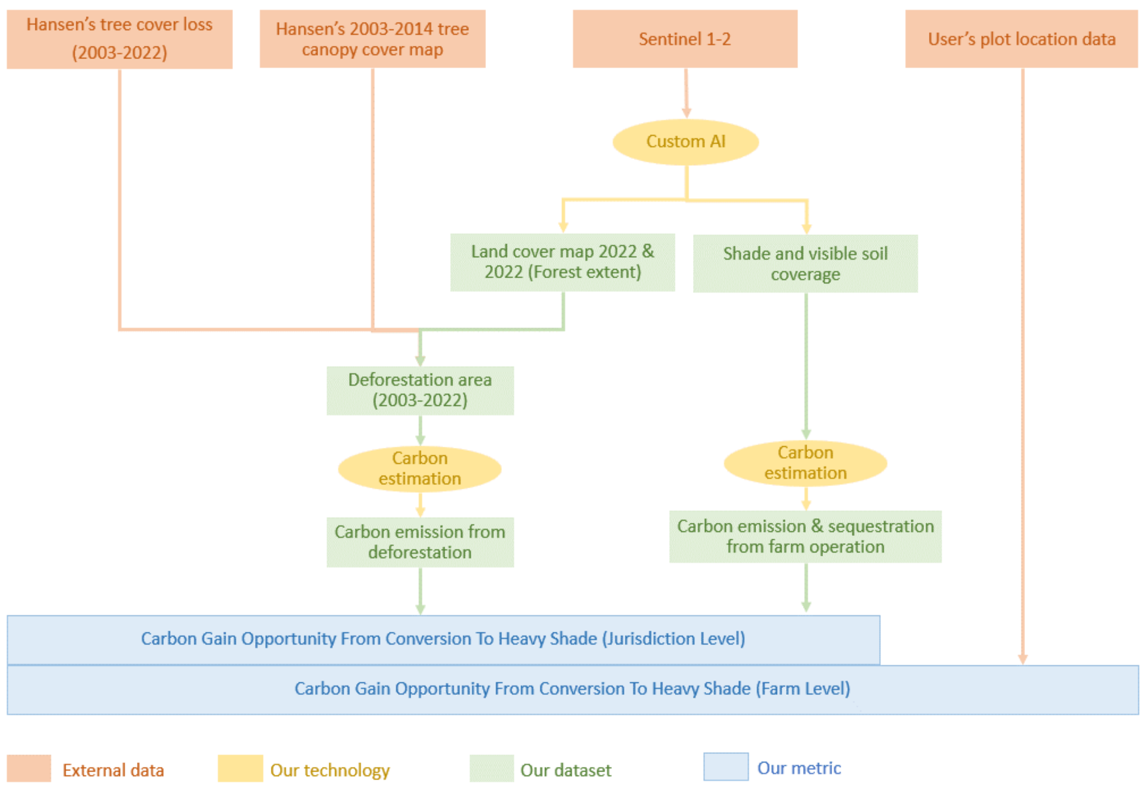
- **By farm unit:** The count of number of farms under each shade level.
  A farm level shade level is assigned to each farm. The shade level assigned to the farm corresponds to the shade level covering the largest farm area.

3.2 Metric: Potential Carbon Gain from Conversion To Heavy Shade.

This metric is used to benchmark (1) jurisdiction level, or (2) farm level carbon gain opportunity from agroforestry intervention, by calculating potential carbon gain if the current coffee shade system is fully converted into heavy shade (>15% canopy cover).

For both jurisdiction level and farm level, the results are computed and displayed in two ways:

- **Absolute value:** Potential carbon gain, in tons of CO2 equivalent.
- **Relative value:** Potential carbon gain, divided by coffee area (tons of CO2 equivalent per hectare).



3.2.1 Carbon emissions from deforestation

Deforestation area for the last 20 years (2003-2022) is calculated from Hansen Forest lost data (see section III.5. External datasets), which was defined as a stand-replacement disturbance or the complete removal of tree cover canopy at the Landsat pixel scale (Hansen et al., 2013).

The Hansen Forest lost data is then overlayed with the following tree canopy cover and forest maps to remove non-forest tree loss:

- Hansen's 2003-2014 tree canopy cover map where canopy cover is >30% (see section III.5. External datasets).

- 2020 and 2022 forest extent from the 2020 and 2022 land cover maps (see section III.1. Land cover).

The removal of non-forest tree loss creates a dataset of forest loss within the last 20 years (2003-2022). Carbon emission from deforestation in the last 20 years is calculated as following:

- Compute zonal statistics to calculate the area of forest lost per year in the last 20 years, within the selected jurisdictions (or farm units)
- Sum all the areas of forest lost harmonized by the age of the event (van Tonder & Hillier, 2014).

$$\sum_{y \in years} \frac{area_y}{current_{year} - y + 1}$$

- *Above ground biomass*: Multiply *area* by the biomass level based on the level found in *biomass per land cover.*

Land cover	Biomass (ton of dry matter/ha)	Country
Natural forest: Tropical dry forest	130	Vietnam

This level will differ from one country to another, and in the future, this level might be different for different forest types.

- *CO2 equivalent*: Compute the carbon stock by multiplying the biomass by carbon fraction (CF) of 0.5 and converted to total CO2e by multiplying by the ratio of the molecular weight of CO2 to that of carbon - 44/12 (IPCC 2006).

3.2.2 Carbon emission and sequestration from farm operation:

The number of hectares under each shade level (see section II. 3.1.) within the selected jurisdictions (or farm units) is calculated with zonal statistics.

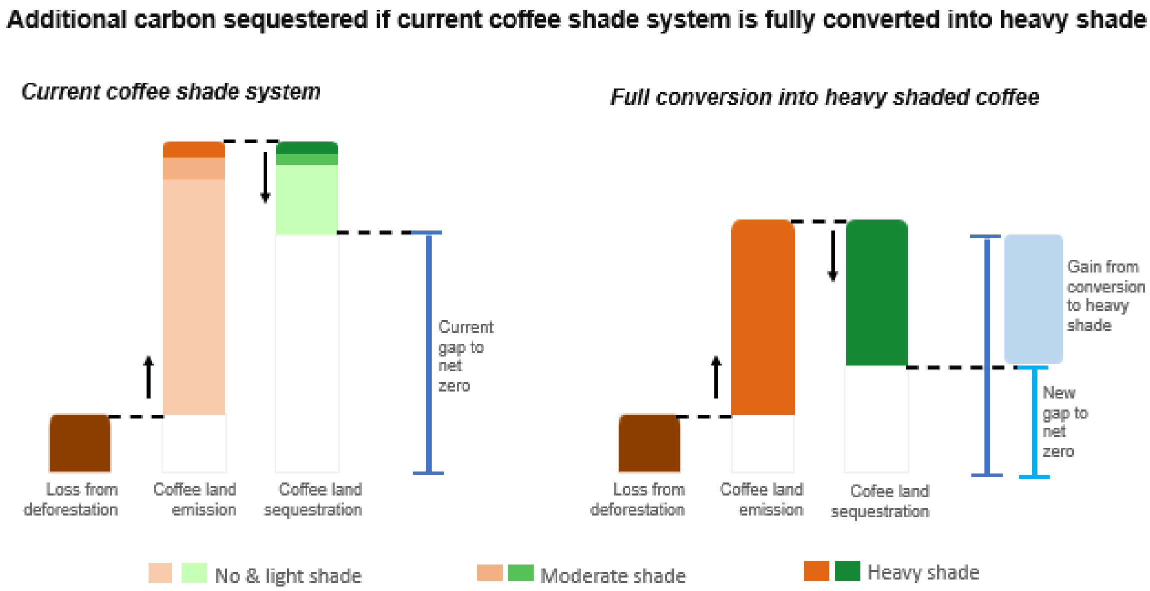
The area under each shade level is then multiplied by the emission and the sequestration factors based on Kuit et al. (2020), to arrive at the emission and sequestration value for each shade level.

The table below lists the emission and the sequestration factors, by shade level:

	No shade (0-5 %)	Light shade (5-10%)	Moderate shade (10-15%)	Heavy shade (>15%)	Country
Emission (Ton CO2e/ha)	3.85	3.85	2.65	2.58	Vietnam
Sequestration (Ton CO2e/ha)	1.14	1.14	1.42	1.93	Vietnam

3.2.3 Carbon gain opportunity from conversion to heavy shade

Carbon gain opportunity is defined as the amount of additional carbon sequestered if the current coffee shade system is fully converted into heavy shade (>15% canopy cover).



First, the carbon balance, or gap to net zero, is calculated for the selected jurisdictions (or farm units) and is then aggregated by summing the different carbon components at the aggregation level selected by the user. The carbon components are:

- Carbon emission from deforestation.
- Carbon emission & sequestration from farm operation, **per shade level**.

The formula used for calculation of the carbon balance, or gap to net zero, is:

Gap to net zero = Emission from deforestation

+ Sum of all shade levels' emission from farm operation

– Sum of all shade levels' sequestration from farm operation

This **Gap to net zero** computation is done for 2 scenarios:

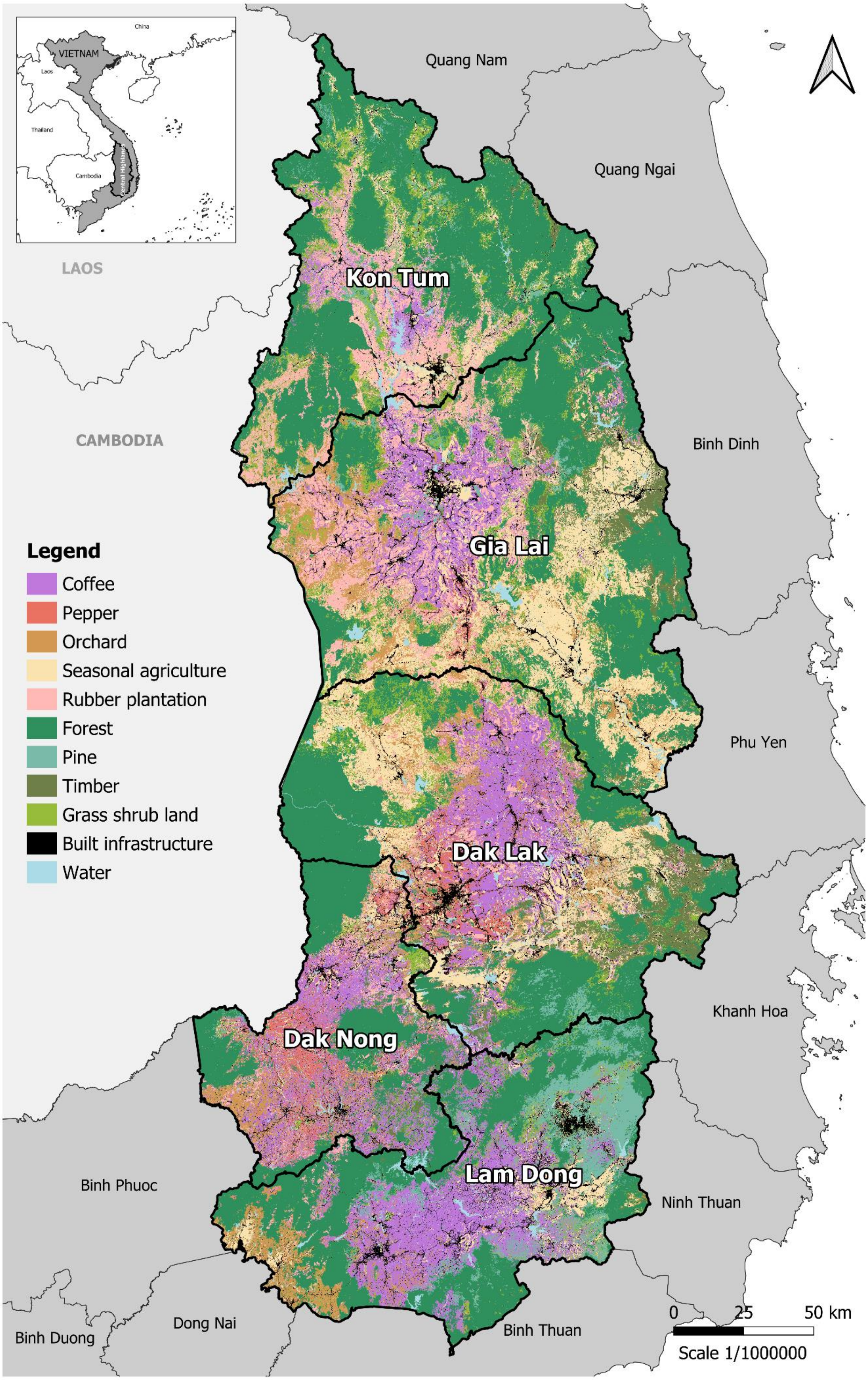
- **Scenario 1 - Current:** The original distribution of shade level as measured Metric: Shade tree coverage (section II. 3.1.)
- **Scenario 2 - Full conversion into heavy shaded coffee: the** shade level being set to Heavy shade (>15%) for all the coffee area.

The potential carbon gain is the difference between the gap of scenario 1 and the gap of scenario 2.

III THE DATASETS

1 LAND COVER

A key innovation of Terra-i+ lies in its ability to provide high-resolution and precise land cover maps, enhancing our understanding of crop distribution in specific landscapes and their impact on forests. The methodology outlined below was employed to produce accurate land cover data for the years 2014, 2020, and 2022. This dataset will be updated on an annual basis and is illustrated in Figure 2.



VIETNAM
China
Laos
Thailand
Cambodia
Quang Nam
Quang Ngai
LAOS
Kon Tum
CAMBODIA
Binh Dinh
Gia Lai
Legend
Coffee
Pepper
Orchard
Seasonal agriculture
Rubber plantation
Forest
Pine
Timber
Grass shrub land
Built infrastructure
Water
Phu Yen
Dak Lak
Khanh Hoa
Dak Nong
Lam Dong
Binh Phuoc
Ninh Thuan
Dong Nai
Binh Duong
Binh Thuan
0 25 50 km
Scale 1/1000000

Figure 2 Commodity specific land cover map in Vietnam Central Highland for the year 2022.

The approach to develop such maps can be succinctly summarized in three primary steps, as shown in Figure 3:

1. **Initial assessment:** This phase involves defining the target levels for the map, compiling the initial dataset used for the first crop specific map creation, establishing, and generating the classification model, and procuring and processing the satellite imagery.
2. **Map improvement process:** The subsequent step is based on an iterative approach, involving the collection of additional data to enhance the map. In this stage, maps are iteratively created and evaluated, and more data are gathered through stakeholder engagement and additional calibration data collection through the interpretation of high-resolution imagery.
3. **Map validation:** The final step encompasses the map validation process, where a dedicated dataset is created, based on field visits and interpretation of high-resolution imagery, and utilized to assess the map's quality, leading to the final version of the crop specific map.

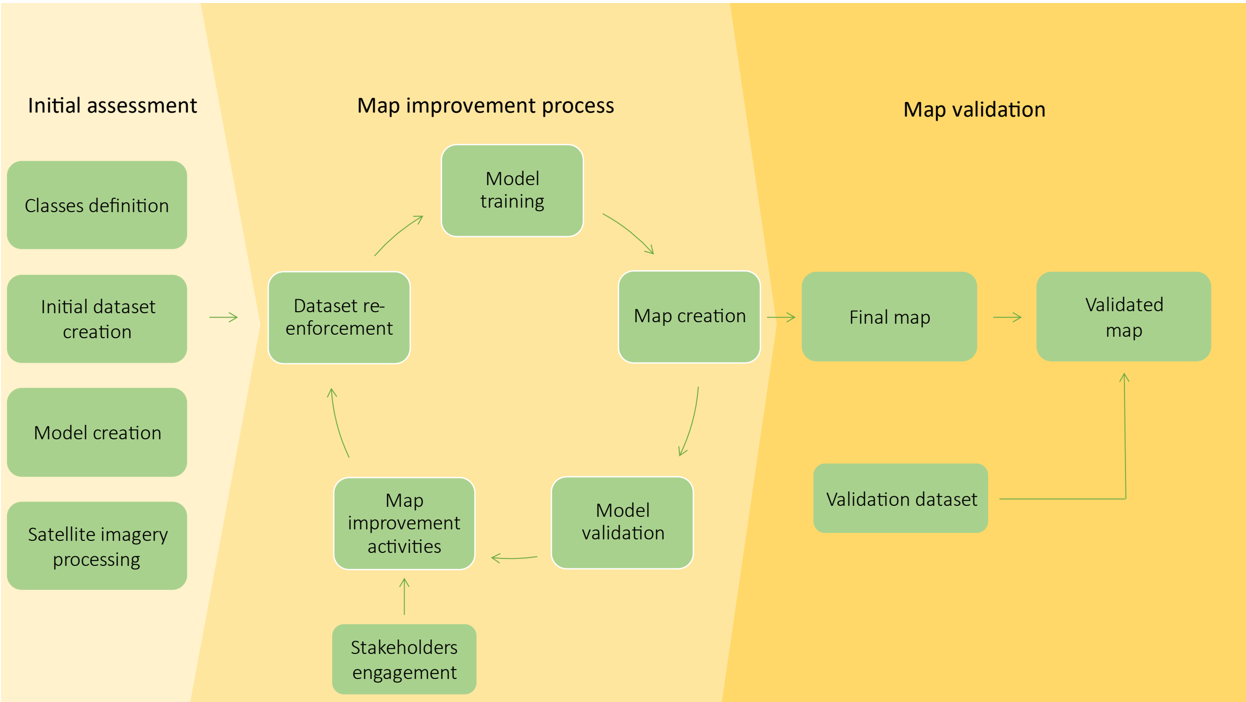


Figure 3 Diagram of the map creation process

1.1 Map categories

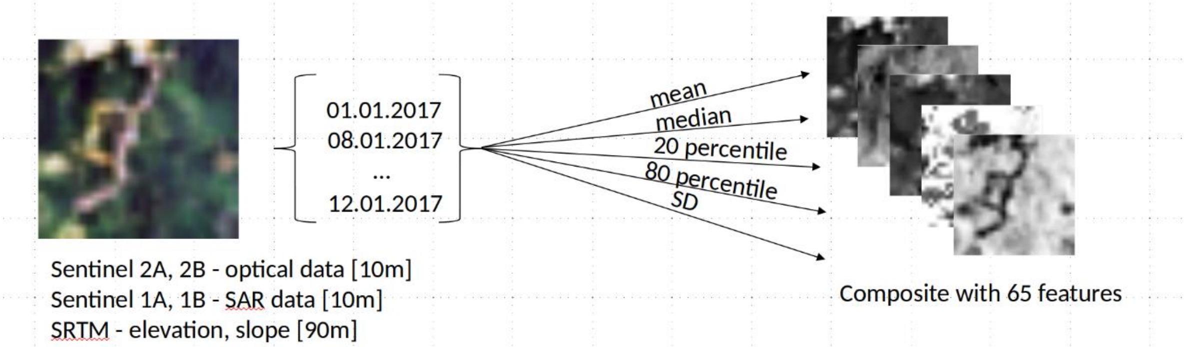
To accurately map land cover categories, two primary criteria must be met. Firstly, it is essential to gather a representative dataset comprising several thousand instances of the category across the designated area. This can be accomplished through the interpretation of high-resolution imagery. The selected categories need therefore to be identified by humans based on the satellite data available. Secondly, the mapped categories must be both exclusive and exhaustive, meaning that each pixel should be assigned to one and only one class, as outlined by Jansen and Di Gregorio, 2000. Considering these criteria, the chosen map categories are detailed in Table 1.

Table 1 Selected and classified land cover.

ASSIGNED CATEGORY ON MAP	DESCRIPTION
COFFEE	Coffee plantation (including intercrop with other fruit tree, and black pepper)
NATURAL FOREST	Forest (with the exclusion of pine tree forest and production forest)
SEASONAL AGRICULTURE	All types of seasonal agriculture
BUILT INFRASTRUCTURE	Human made infrastructure, (house, greenhouse, …)
WATER	River, lake, and artificial water bodies
PINE FOREST	Forest consisting mainly of pine trees
RUBBER PLANTATION	Rubber plantation
GRASSLAND, SHRUBLAND AND BARE SOIL	Mainly Natural land with very low density of tall trees, also includes areas with bare soil and mining areas.
PRODUCTION FOREST	Timber planation, mainly acacia and eucalyptus
ORCHARD	Any orchard not intercropped with coffee or black pepper.
BLACK PEPPER	Black pepper plantation

1.2 Satellite imageries

The model's input data is a combination of satellite imagery from Sentinel 1, Sentinel 2, and SRTM missions, processed using the Google Earth Engine platform. Sentinel 1 imagery undergoes pre-processing, including orbit correction and noise removal. For Sentinel 2, cloud-free imagery is selected, and a quality mask is applied to further eliminate clouds. Each spectral band is then processed by calculating annual percentiles, mean, and standard deviation, resulting in a composite of 65 features. The inclusion of percentile and standard deviation helps retain intra-annual temporal information for improved land cover prediction. Figure 1 provides a summary of the entire process.



1.3 Reference data collection

A field trip was initially conducted in the study area to identify representative occurrences of each category in Table 1. An iterative sampling approach based on human interpretation of high-resolution satellite imagery was then implemented. Maps generated at a given iteration are the basis for subsequent rounds of sampling, focusing on areas with lower system performance. This process continued until satisfactory quality was achieved. The image interpreter primarily utilized Google imagery through Google Earth Pro, examining time series to enhance crop type identification and verify imagery dates. Planet data was used for additional verification of land cover changes. Google Street Map imagery was also included when available. The current map results from 25 sampling/mapping iterations, utilizing a dataset of 70,000 GPS locations in the Central Highlands as shown in Figure 4.

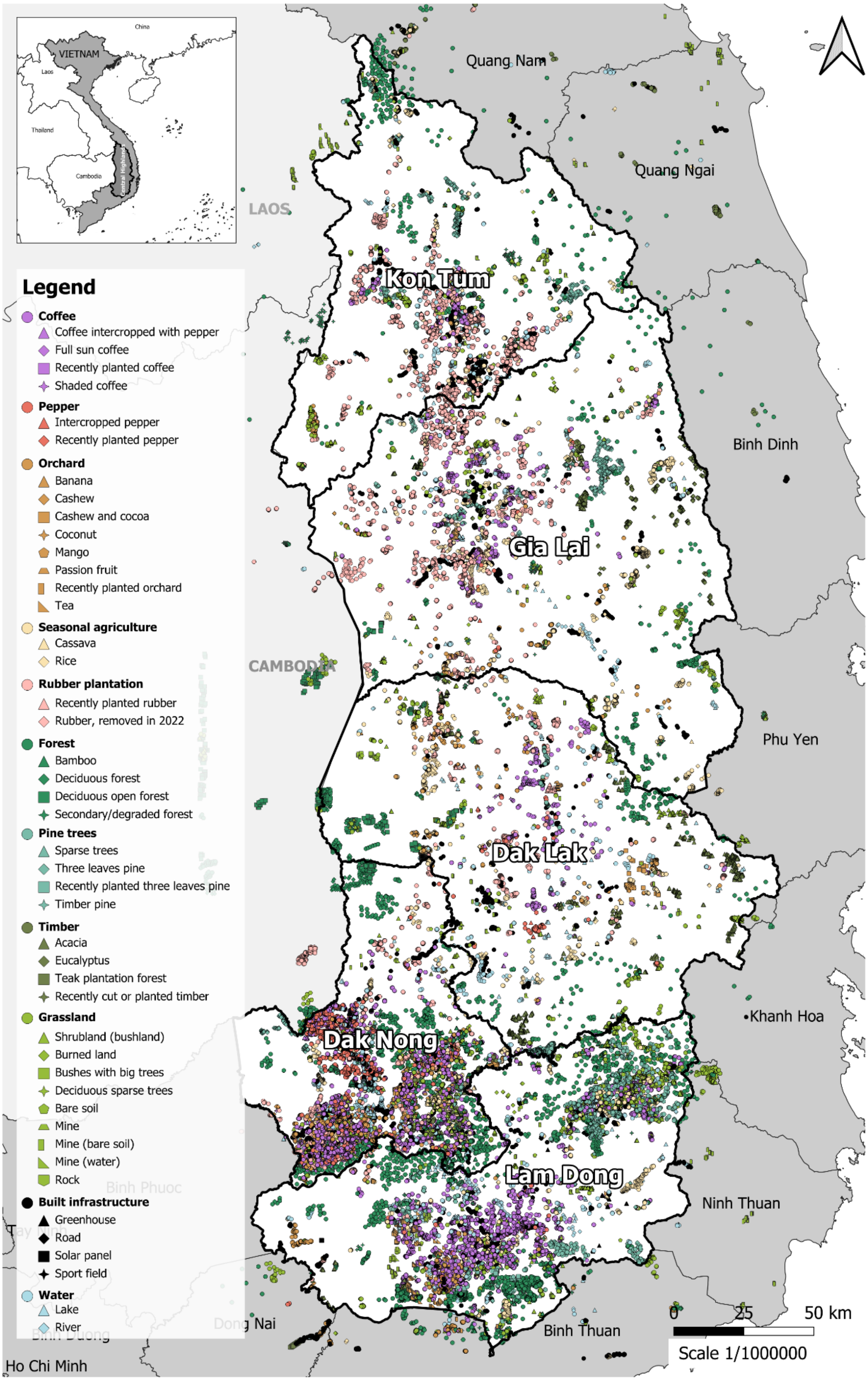


Figure 4 GPS location around the area of interest: Google map.

1.4 Models

Our classification approach is based on the use of an ensemble of deep convolutional neural network as shown in Figure 5. In this work, we trained five models with slightly different versions of the training dataset based on a k-fold splitting strategy. Each model sees 80% (4 of the 5 data split) of the data during training and uses the 20% left to pick the best epoch. A map was generated for each of the five models, and for each pixel, the land cover value was chosen as the value appearing the most often on the 5 maps. As the model ensemble assigns a land cover at pixel level, the final map has the same resolution as the input data. Around 10X10m for sentinel imagery.

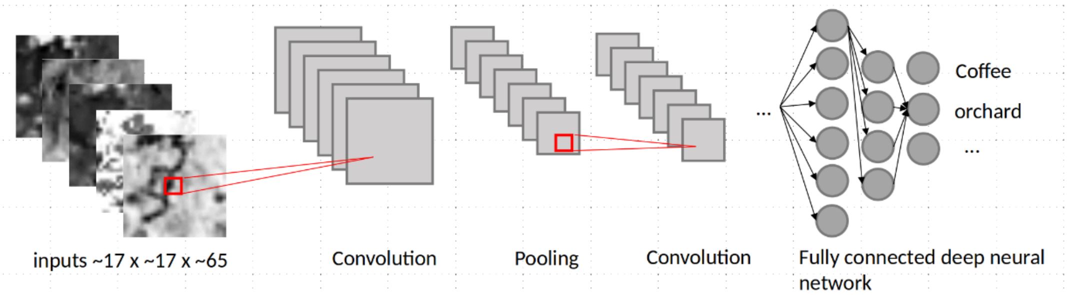


Figure 5 Approximate architecture of the model. The model receives a small window and output the categories of the central pixel.

1.5 Validation

The map validation process followed a rigorous protocol inspired by methodologies outlined in (Olofsson et al., 2014), with adjustments proposed by (Stehman, 2014). The validation method was based on high-resolution imagery in two phases. In the initial phase, image interpreters reviewed points that were sampled following a stratified random sampling. They assigned land covers to each point, considering a 10-meter buffer to account for spatial variation and satellite resolution. Uncertain points, for which interpreters did not agree, were then identified. The second phase involved a detailed revie w exercise of each uncertain point by all the interpreters as a group. Each point was finally given a class based on votes. The review aimed for a 2.5% margin of error at a 95% confidence level for overall map accuracy. Based on these points, a confusion matrix was generated and normalized to compute key performance metrics, including overall accuracy, precision, and recall. Confidence intervals for validation metrics and crop areas were calculated considering the approximate distribution of each class.

COVER	*PRECISION*	*RECALL*	*AREA [HA]*	*AREA RATIO*
Coffee	91% ± 8%	82% ± 10%	901,777 ± 14%	16.20%
Forest	88% ± 9%	90% ± 3%	2,020,796 ± 10%	36.30%
Seasonal agriculture	81% ± 12%	82% ± 7%	845,073 ± 15%	15.20%
Built infrastructure	81% ± 12%	98% ± 4%	173,163 ± 15%	3.10%
Water	79% ± 12%	96% ± 7%	128,309 ± 16%	2.30%
Pine	77% ± 12%	96% ± 5%	155,227 ± 16%	2.80%
Rubber plantation	70% ± 14%	85% ± 14%	313,198 ± 23%	5.60%
Grass shrubland	68% ± 15%	30% ± 11%	606,546 ± 34%	10.90%

Timber	62% ± 14%	76% ± 27%	116,289 ± 40%	2.10%
Orchard	46% ± 15%	70% ± 19%	303,315 ± 35%	5.50%
Overall Accuracy	**80% ± 5%**			

1.6 2014 land cover map

Several methodological adjustments were necessary to produce the 2014 land cover map. Because Sentinel 2 had not been launched by that time, the map relies on Landsat 8 data resampled at a 15m resolution. Initially, Landsat images were resampled at a 60m resolution, and a convolutional neural network was trained to upscale them to their original 30 meters resolution using Landsat's panchromatic information. This model was then applied to the original data, resulting in Landsat images with a 15m resolution. Additionally, due to the limited availability of high-resolution data in 2014, the mapped categories were limited to coffee, forests, and others land covers. Aside from these adjustments, the map generation process followed the same approach used for the more recent years.

2 SHADE AND SOIL MAP

The second major innovation of Terra-i+ is its capability to deliver precise assessments of shade levels within coffee systems at high resolution, thereby enriching our comprehension of agroforestry distribution in targeted landscapes and its influence on carbon sequestration. The methodology described below was applied to generate accurate shade level data for the years 2020 and 2022, with plans for annual updates. Subsequently, we detail the approach used to identify the density of shade trees within specific agricultural units. To ensure result consistency, a close alignment was maintained between the input data and the models used in crafting land cover maps.

2.1 Reference data collection

For the precise identification of shade levels within coffee systems, a meticulous reference dataset was crafted. The process starts with the identification of shade trees within cells ranging from 0.07 to 3.4 square kilometers, covering a total area of 25.7 square kilometers. Employing high-resolution imagery, an initial automated analysis identified shade trees, followed by a step of manual correction by human experts to refine and ensure the high quality of the reference data. Post-refinement, the high-resolution dataset underwent down-sampling to a resolution of 10x10 meters as shown in Figure 6. This down-sampling allowed for the representation of shade tree density within each Sentinel pixel, ensuring compatibility with the model's operational scale. The integration of both automated and manual approaches in dataset creation aimed to enhance the accuracy and reliability of the reference dataset, establishing a robust foundation for training the model to effectively discern shade levels within diverse landscapes of coffee systems.

Google imagery High resolution mask 10X10m down sampled

Figure 6 Reference data creation for shade tree density identification.

2.2 Models

The model used for assessing shade levels in coffee systems follows a similar structure to the one utilized for crop-specific map classification. The model maintains consistency in both inputs and convolutional structure than the land cover classification framework. Nevertheless, the task involves a regression problem, diverging from the discrete nature of the classification model. In this context, the model's output is tailored to address shade level density, presenting a continuous interval prediction. The resulting model structure enables the system to effectively capture and quantify nuanced variations in shade levels, providing valuable insights into the intricate dynamics of agroforestry systems.

2.3 Validation

A test dataset constituting 15% of the entire reference data and randomly selected, was used to assess the model's generalization capabilities. The points in the dataset were then classified into low and high intercropping levels using a threshold of 20% shade tree canopy cover. Subsequently, the model is applied to this unseen dataset, and the results are compared with the reference data. The overall accuracy, along with the user and producer accuracy of the shade level, is then computed. The results are shown in Table 2, presenting an overall accuracy of 86% but a relatively lower precision and recall for the highly shaded systems, indicating that our method tends to slightly underestimate the shade level.

Table 2 Shade level accuracy assessment

	PRECISION/USER ACCURACY	RECALL/PRODUCER ACCURACY
HIGH SHADE	69%	65%
LOW SHADE	90%	91%
GLOBAL ACCURACY	**86%**	

3 DEFORESTATION MONITORING SYSTEM

Terra-i is a near real time monitoring system based on MODIS imagery for natural vegetation loss detection resulting from human activities. Terra-i was first implemented in Latin America in 2012, producing updates every 16 days with the current coverage for the whole of Latin America and the tropics. Since June 2012, Terra-i data have been available free of charge for download on www.terra-i.org.

With the aim to detect forest cover loss at small scale to meet local needs, the latest version of Terra-i (Terra-i v3.0) was developed, using Sentinel 1 Synthetic Aperture Radar (SAR) images with a spatial resolution of 10m. It was first piloted in Di Linh district of Lam Dong province in 2017, and is currently producing frequent updates in Tuong Duong and Ky Son district, Nghe An province and five provinces of the Central Highlands.

Terra-i objective is to detect forest loss using Sentinel 1 Synthetic Aperture Radar (SAR) images provided by European Space Agency (ESA). SAR data offers many advantages as it is little affected by cloud cover or a lack of illumination and thus can operate under all weather conditions as well as during day or night. Sentinel 1 mission is currently composed of two satellites Sentinel 1A and 1B. By combining the measurements from both satellites, a given site is visited every 6 days.

Sentinel-1 sensor offers two measurements (bands) for each observed pixel, including the magnitude of the waves that were transmitted vertically polarized and received vertically polarized (VV) and the magnitude of the waves that were transmitted vertically polarized and received horizontally polarized (VH) after being backscattered.

Nonetheless, SAR data also has limitations when it comes to monitor vegetation in general and forest in particular. For instance, the images present speckles resulting from the high heterogeneity of the structure of forest canopy. SAR data are also limited in comparison to optical data when it comes to monitor the status of vegetation given the relatively reduced number of bands available and the inexistent interaction between the vegetation chlorophyll and the SAR electromagnetic waves. To overcome the issues due to the high heterogeneity of the forest canopy we firstly mask all the non-forested areas using the land cover map from 2020.

The second step of this methodology is to identify abnormal changes between pairs of consecutive images within forested areas. Indeed, although the forest canopy texture presents a high heterogeneity, the canopy structure at a given site does not change drastically in 6 days. To do so, Terra-i v3.0 is based on a robust three-step change detection method. Initially, it defines the normal range of differences between two consecutive images by utilizing the radar cross-polarization bands VV and VH. Subsequently, the multivariate normal distribution of deviation from one image to the next is determined for various types of forests, such as broad leaf forests and deciduous forests. This involves calculating the mean, standard deviation for each polarization, and assessing the correlation between the two polarizations, VV and VH. Finally, anormal deviations from one image to the next are identified and categorized as forest loss, pinpointing deforestation events with accuracy. This method allows Terra-i v3.0 to effectively identify and map changes in forest cover, facilitating a comprehensive understanding of deforestation dynamics in near real time as shown in Figure 7.

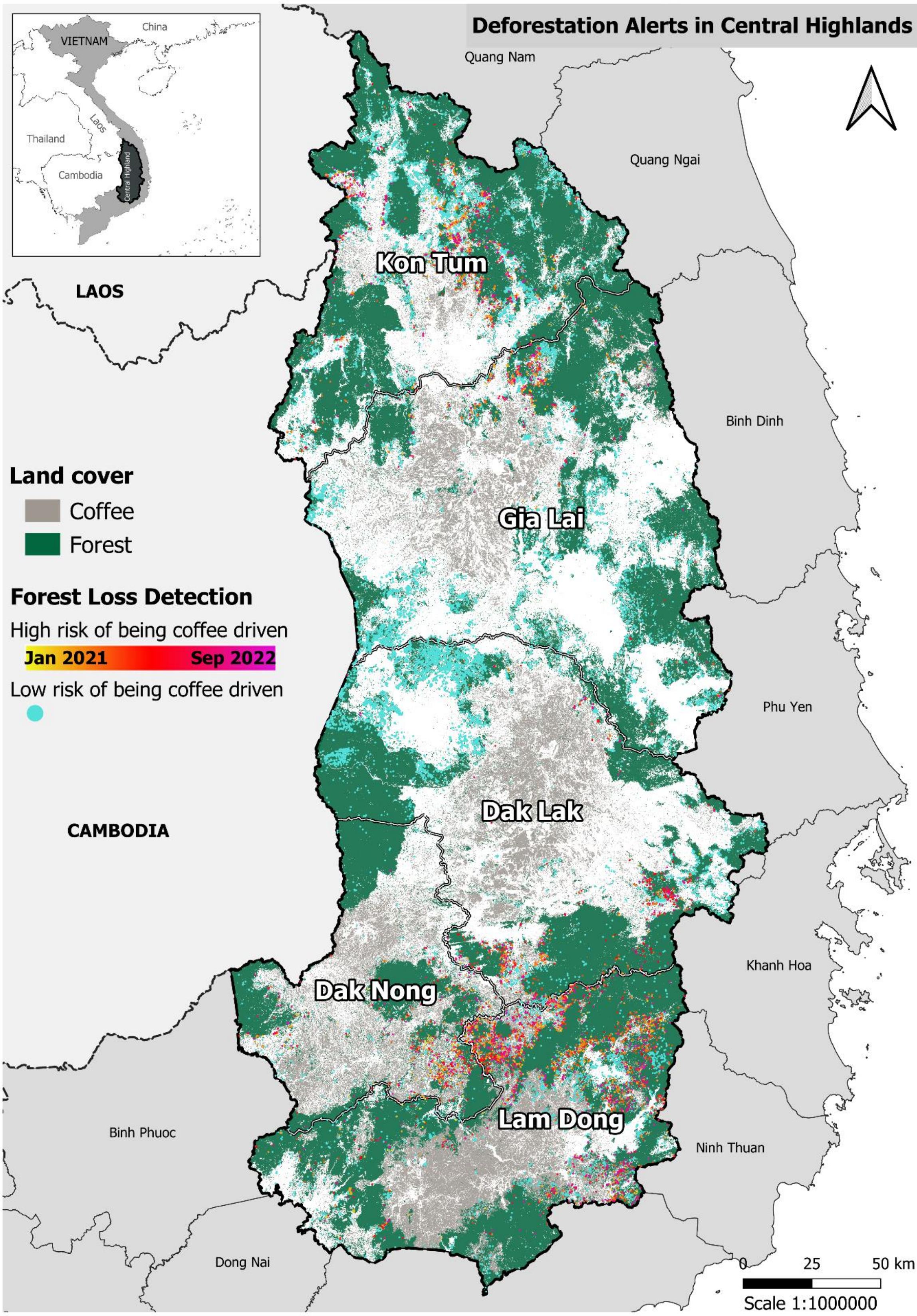


Figure 7 Deforestation events in the Central Highlands, Vietnam (2021 - 2022)

4 CLIMATE SUITABILITY

Our model shows the potential areas for coffee production along the highlands range. Four agro-climatic zones were defined, differ by the annual rainfall amounts and the temperatures: Remain/become suitable; High adaptation need; Highest adaptation need; Remain/become unsuitable.

A climate projection is the simulated response of the climate system under a future scenario of emissions or concentration of greenhouse gas (GHG), usually derived from global climate models (GCM). A GCM is a representation of the climate system based on the physical, chemical, and biological properties of its components, their interactions, and feedback processes. Climate projections are contingent on the emission scenario used, which in turn is based on assumptions concerning future socio-economic and technological developments in A Shared Socio-Economic Pathways (SSP). GCM outcomes have a coarse resolution of 100 or 200 km, which is not practical to assess agricultural landscapes. Therefore, we use downscaled climate projections. The key assumptions of this approach are that changes in climate only vary over large distances and that the relationship between variables in the baseline is maintained in the future.

Emission scenarios are a plausible representation of the future development of GHGs. Optimistic emissions scenarios assume that net carbon emissions become zero in the near future (2021-2040), while in the pessimistic SSP 570 scenario, GHG emissions keep growing, resulting in extreme warming. Several publications show that in this scenario coffee would struggle to survive. In this study, we used SSP 370, because it is an appropriate option (intermediate scenario) to guide adaptation.

To determine zones with different degrees of climate impact, we modelled changes in bioclimatic suitability for coffee under present conditions and those in the period between 2021 and 2040 (which we approximated to 2030) using a machine learning classification model.

- First, we assembled a database of locations where coffee is currently grown.
- Second, we interpolated monthly climate means of the 1970–2000 period onto a 0.5 arcminute grid, which were downloaded from the WorldClim database (Hijmans et al., 2017), representing our current baseline climate. They were used to calculate 19 bioclimatic variables commonly used in modelling of crop suitability (Nix, 1986).
- Third, applying Random Forests in unsupervised variations to detect biologically significant clusters of coffee suitability within the occurrence data using the bioclimatic variables as features. These clusters can be interpreted as different climate zones that allow growing coffee, but under different climate conditions.
- Fourth, using all bioclimatic variables, Random Forest clusters were trained to distinguish between suitable areas (falling into one of the suitable climatic zones) and unsuitable areas for coffee. Clusters were applied to climate data from the 19 climate scenarios of the 2030 periods on the basis of equally likely GCM outputs . This resulted in 10 distinct suitability maps that were averaged to obtain one single map for each period (2030).
- Finally, recommendation domains were defined according to the quality of change between climatic zones under current conditions and under future conditions in each of the 10 GCM projections.

As any future outlook, our model has a considerable degree of uncertainty and should be considered only as a projection, not a prediction. The uncertainty in our model comes from the emission scenarios, the climate models, and the crop model. We used 10 global climate models as equally valid projections of future climate. These models show a high level of agreement on the increase of temperature, but they disagree on the regional and seasonal distribution of rainfall. Therefore, the resulting impact gradient is largely influenced by the temperature increase, while the disagreement on rainfall is masked. However, an increase in temperature means a greater demand of water for agriculture. Lastly, our model is an "all other things equal" model that only considered a change in climate. Our statistical approach is designed to avoid overfitting and it also deliberately includes marginal locations for coffee. This should be considered "friendly" uncertainty because it means that through guided adaptation, the worst impacts will be avoidable.

5 External datasets

Global Forest Change: The results of the time-series analysis of Landsat images. The dataset depicts the characterization of forest extent and transition from a forested to a non-forested state, within the time frame spanning from 2000 to 2022. (Hansen et al. 2013).

IV References

Hansen, M. C., Potapov, P. V, Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V, Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. R. G. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. Science, 342(6160), 850–853. https://doi.org/10.1126/science.1244693. Data available online from: http://earthenginepartners.appspot.com/science-2013-global-forest.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 25.

IPCC (2006). 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Prepared by the National Greenhouse Gas Inventories Programme, Eggleston H.S., Buendia L., Miwa K., Ngara T. and Tanabe K. (eds). Published: IGES, Japan.

Kuit, M., Jansen, D.M., Tijdink, N. (2020). Scaling up sustainable Robusta coffee production in Vietnam: Reducing carbon footprints while improving farm profitability - Full technical report. https://www.idhsustainabletrade.com/uploaded/2021/03/Scaling-up-Sustainable-Robusta-Coffee-Production-in-Vietnam-full-tech-report_March-102021.pdf. (Summary version available at https://pdf.usaid.gov/pdf_docs/PA00ZF26.pdf)

Nix, H.A. (1986) A Biogeographic Analysis of Australian Elapid Snakes. In: Longmore, R., Ed., Atlas of Elapid Snakes of Australia. Australian Flora and Fauna Series No. 7, Australian Government Publishing Service, Canberra, 4-15.

Turubanova S., Potapov P., Tyukavina, A., and Hansen M. (2018) Ongoing primary forest loss in Brazil, Democratic Republic of the Congo, and Indonesia. Environmental Research Letters https://doi.org/10.1088/1748-9326/aacd1c van Tonder, C., & Hillier, J. (2014). *Technical Documentation for the online Cool Farm Tool*. 34. https://coolfarmtool.org/