A ground truthed assessment of University of Maryland forest change dataset in four tropical regions against a deep learning generated dataset

Grace Gee*, Eugene Wang*

Enveritas, 24 Innis Lane, Old Greenwich, Connecticut, U.S.A.

*Equal contribution

24, December 2021

Correspondence: grace@enveritas.org

Keywords: Deforestation; monitoring; remote sensing; deep learning; ground truthing; tropics

ABSTRACT

Introduction

Governments, corporations and civil society have a strong motivation to protect forests and take action against deforestation, particularly if the deforestation has occurred as a result of exportable commercial commodities such as coffee or cocoa. Corporations also need accurate deforestation measurement as they seek to assure deforestation-free supply chains and set science-based GHG targets.

However, there have been significant limitations in the world's ability to accurately measure deforestation. The main deforestation model utilized by major global environmental organizations is generated by Global Forest Watch (GFW). The GFW model has the following major shortcomings:

- It is unable to differentiate commercial tree plantations from forest
- It cannot penetrate cloud cover, which frequently obscures optical satellite images in the tropics
- It has difficulty detecting small-scale deforestation, such as encroachment for firewood, charcoal, and farming activities
- It lacks feedback from ground truthing, i.e., validating computer model accuracy by conducting on the ground observations
- It does not account for reforestation events; a "tile" is no longer monitored once it has been classified as deforested

Since GFW was released in 2012, there have been several significant advances in technology, particularly with regard to satellites and deep learning methods.

This project piloted a new approach, which takes advantage of these recent improvements in technology, and aims to offer an improved deforestation model.

In this white paper, we provide a description of the Enveritas deep learning approach and how it combines deep learning, satellite imagery and innovative ground truthing methods to detect small-scale deforestation and remove planations from forests. We then validate Enveritas and GFW approaches across four tropical regions -- Southern Ethiopia coffee regions, Dak Nong (Vietnam), Côte d'Ivoire, and Aceh and North Sumatra (Indonesia).

Summary of Findings

Enveritas found lower levels of deforestation than GFW in all four regions:

- Southern Ethiopia coffee regions: 48% lower
- Dak Nong (Vietnam): 40% lower
- Côte d'Ivoire: 80% lower

- Aceh & North Sumatra (Indonesia): 46% lower

The main cause of GFW's overestimation is the inability to remove plantation tree cover loss from deforestation accounting. This error is highly variable and dependent on crop types and agronomy practices. For instance, in Aceh and North Sumatra where the dominant plantations (e.g., palm and rubber) exhibit high tree cover, we find that 55% of GFW's 2019 tree cover loss events is due to plantation removal. In contrast, Dak Nong plantations tend to be lower in tree cover and only 27% of GFW's 2019 tree cover loss events is due to plantation removal. In differentiating forest from plantations, the GFW tree cover map achieved ground truthed precision of 45-75% whereas the Enveritas model achieves precision of 80-94%.

Within forest areas only, GFW misclassifies 12-32% deforestation events compared to the Enveritas model.

In addition, GFW has difficulties in detecting small-scale deforestation. Across all four regions, the average Enveritas detected deforestation cluster size is smaller than that detected by GFW. In Ethiopia, where small encroachment is the paradigm of deforestation, the average Enveritas deforestation cluster size is less than a fifth of the average GFW tree cover loss cluster size.

Implications

GFW is the only public global deforestation dataset and therefore it is frequently used in civic, corporate, and government discussions related to deforestation and climate change.

This project shows GFW's high variation of deforestation accuracy and tendency to overestimate the area deforested makes it unsuitable for tracking deforestation across time and region. At the same time, their limitations in detecting deforestation events within forest areas, especially small-scale ones, would hamper timely intervention.

The Enveritas approach opens new possibilities to couple region-specific models with ground truthing to have a more accurate depiction of deforestation.

Funding

This work was supported by HereWeGrow.

BACKGROUND

How GFW works

GFW uses Landsat imagery at 30m resolution to build 2 models - a 2000 Tree cover % model which takes in 2000 Landsat data and outputs tree cover % of each 30m pixel and a tree cover loss model which takes in a specific year's Landsat data and outputs whether the 30m pixel has been converted to bare land.

GFW defines tree cover as "all trees and vegetation taller than 5m regardless of whether they're part of a natural forest or have been planted by humans for agricultural production." Tree cover loss is defined as "complete removal or mortality of tree cover canopy at the Landsat pixel scale". Tree cover loss does not necessarily equate to "deforestation" and can result from a variety of factors, including mechanical harvesting, fire, disease, or storm damage.

The GFW platform operates on 2 assumptions. First, it assumes that when a pixel is converted to bare land, the tree cover % lost reported is the % measured in 2000, not in the most recent year. Secondly, a pixel can only lose all its tree cover once.

GFW limitations

Coffee and cocoa producing areas are located in the tropics and commonly exhibit patterns of small-scale deforestation.

GFW has difficulty detecting such small-scale deforestation in the tropics due to data scarcity caused by their reliance on optical satellites which are obscured by cloud cover more than 75% of the time in tropical regions, 75% of Landsat images had too much cloud cover to be useable in this study[1] and 2). Mitchard et al. reports 94% of deforestation went undetected in a region analyzed in Ghana because "much forest loss occurs at a very small scale."[2]

The GFW models use decision trees[3], which are more sensitive to noise or missing data in the multispectral bands than deep learning methods. Figure 1 shows two examples of how GFW's decision trees' cutoffs can lead to inaccurate predictions.

The GFW model assumptions are not sufficient for monitoring deforestation and reforestation. Tree cover % doesn't differentiate between plantations and forest. In areas with tree-like

---

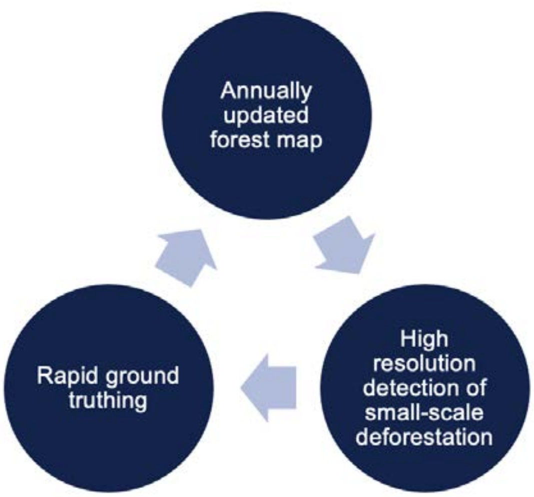
[1] We counted the number of 2018 Landsat 8 images (GFW's 2018 source images) in the 4 target regions and calculated the percent of images having less than 20% cloud cover.
[2] Mitchard, Edward, et al. "Assessment of the accuracy of University of Maryland (Hansen et al.) Forest Loss Data in 2 ICF project areas–component of a project that tested an ICF indicator methodology." *URL: https://ecometrica.com* (2015).
[3] Hansen, Matthew C., et al. "High-resolution global maps of 21st-century forest cover change." *science* 342.6160 (2013): 850-853.

plantations (e.g. rubber, cocoa, palm), this results in overstated deforestation measurement since plantations are more likely to register land use changes. Because a pixel can only lose all its tree cover once, GFW ceases to monitor reforested areas for subsequent deforestation.

The GFW validation method can be improved in 3 areas. Firstly, GFW results are not ground truthed. Validation is conducted by manually interpreting high resolution satellite imagery of 1,500 120m blocks for tree cover loss and gain. Secondly, as a result of using larger 120m blocks instead of 30m (their output resolution size), GFW's evaluation is focused on detecting larger, more conspicuous forms of deforestation. This is supported by their conclusion that their result "illustrates a robust product at the 120m pixel scale". Thirdly, analysis is conducted over a cumulative period of 2000 to 2012, not for an individual year.

How Enveritas deforestation accounting works

The Enveritas approach combines annually updated forest maps with the unique ability to monitor small-scale deforestation and conduct rapid ground truthing. In contrast to GFW, Enveritas makes annually updated forest maps and ground truthing core parts of deforestation monitoring.



We define forest as land spanning at least 10m x 10m with trees higher than 5 meters and a canopy cover of more than 10 percent. It does not include land that is predominantly under agricultural or urban land use. This definition is an expanded version of the FAO definition[4]. The only modification is we reduced the minimum area a forest must span from 0.5 hectares to 10m x 10m since our model can detect forest on a more granular level. Deforestation is defined as the

[4] http://www.fao.org/3/ap862e/ap862e00.pdf

conversion of forest to other land use or the permanent reduction of the tree canopy cover below the minimum 10 percent threshold.

Enveritas models use multiple high resolution optical satellite imagery sources combined with radar imagery to address the issue of frequent cloud cover in the tropics, increase the model's sensitivity to smaller scale deforestation, and output results on a high resolution basis. Radar satellites can penetrate cloud cover using microwaves, and have been shown to be useful for detecting canopy structure. The combination of multiple optical satellites provides us with higher frequency data and smooths out noise.

Enveritas builds 2 models - 1) a localized forest differentiator model which is a deep learning model capable of using diverse imagery sources at various resolutions to accurately differentiate forest from other land types in a localized context; the model is run annually to account for reforestation and 2) a deforestation model which is an ensemble model that takes in the same diverse inputs as the first model but predicts whether a forest pixel has been deforested in a given year.

We output results on a 10m basis which means that our pixels and accounting will be more granular. Figure 2 shows how Enveritas' 10m output resolution results in a sharper forest boundary compared with GFW's 30m output resolution.

The Enveritas approach operates on 2 assumptions. First we monitor only forest pixels instead of tree cover pixels like GFW does. Secondly, each pixel can have the possibility of being reforested and we continue to monitor for repeat deforestation.

We validate our models for 4 regions using visual interpretation (VI) of high resolution satellite imagery with the help of expert agronomists and ground truthing.

Innovation realized through Enveritas deforestation ground truthing methods

Evaluating a deforestation product is difficult due to the number of ground truthed pins required to achieve an acceptable level of accuracy and lack of historical land cover data. Thus many past studies rely solely on visual interpretation to evaluate deforestation products.

Enveritas' ground operation permits ground-truthing to be conducted at scale.

METHODS

Research area

Regions analyzed include 1) Dak Nong, a district of Vietnam, 2) the entire Côte d'Ivoire, 3) coffee-growing regions of Ethiopia, and 4) North Sumatra and Aceh, for a total area of almost 50 million hectares (Figure 3) The areas represent a diversity of coffee and cocoa smallholder landscapes, industrial plantations and natural land types.

Satellite Imagery Data

We use data from multiple optical and radar sources - DigitalGlobe (0.5-1.5m), Sentinel 1 (10m), Sentinel 2 (10-20m), Landsat (30m) and Planet (5m).

For Sentinel 2, we use top-of-atmosphere reflectance data and the QA band to filter cloud cover on an image granule level. On the pixel level, we further apply a custom cloud cover filter. For each year, three sets of imagery are curated, corresponding to quality mosaics which maximize the normalized difference vegetation index (NDVI, using B4 and B8 bands), normalized difference of the SWIR and green band (B11 and B3 bands) as well as a custom formula of B11, B3 and B4 bands. These mosaic criteria are chosen to ensure a highly continuous imagery across the entire region while having the greatest distinction from each other.

For Landsat 8, we use surface reflectance data and apply the cloud filter generated by the CFMASK algorithm on a pixel level in combination with our custom cloud filter. Similar to the Sentinel 2 case, we utilize three sets of imagery for a given year, corresponding to quality mosaics maximizing NDVI, normalized difference of the B6 and B3 bands, as well as a custom formula of B3, B4 and B6 bands.

We compare the mosaic approach of curating optical multispectral imagery to a time series compilation to understand the implications of missing imagery due to extensive cloud cover.

As not all bands within Landsat 8 and Sentinel 2 have a native resolution of 10m, we apply a simple reprojection to those bands without any interpolation. We investigated if interpolation techniques such as bilinear and bicubic introduce visual inconsistencies when compared with high resolution imagery.

For Sentinel 1 SAR, we utilize VV (vertical transmit/ vertical receive) and VH (vertical transmit/ horizontal receive) in both ascending and descending orbital passes. Unlike the optical imageries, radar data tends to be noisy and contains speckles. To smooth the data, we utilize a modified refined lee speckle filter. We use radar to resolve missing imagery issues and we choose to use a time series representation of monthly sar data.

For Planet imagery, we use the normalized analytic biannual imageries, which contains red, green, blue and nir bands at 5m resolution in tropical regions.

For DigitalGlobe high resolution RGB imagery, we use imagery from satellites with ground sampling distances of around 0.5-1.5m. High resolution imagery is best utilized in conjunction with other satellite sources. In particular ,we use the 0.5-1.5m imagery for forest differentiation and for visual inspection validation work.

Label Data

The objective is to have a set of positive training examples of ‘cleanly’ labeled forest pixels and forest loss pixels by cross-referencing various datasets. To create a set of negative non-trivial but representative training examples (non-forest / no forest loss pixels), we choose to randomly sample in the vicinity around the positive pixels as well as generally within the region of interest.

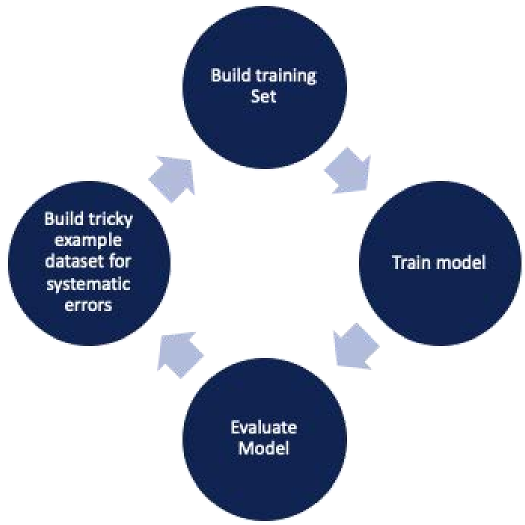
Since the positive training examples are in the minority, the model will be more sensitive to mislabels in the positive than the negative set. Thus we can be more liberal with introducing mislabels to the negative set (since it is randomly curated) and approach the data as a positive-unlabeled set.

For the forest training set, we gather global and regional land cover sources, prioritizing for recent and high resolution datasets which include forest, regional crops, or land cover data.

To aggregate the sources, we cross-referenced the datasets to keep only consistent forest labels. We choose to cross-reference the dataset instead of relying on a single source as they correspond to different years with different levels of granularity. Upon inspection, we also found most of them to be vaguely accurate and many do not tag the entire region we are operating in.

Table 1 contains the data sources used. Figure 4 shows the superior output of the Enveritas forest differentiator model compared to other label data sources used for training in terms of resolution and precision.

After curating the initial training set, we follow an iterative process to further expand the dataset using the iterative method below. The models are not geographically transferable - in the sense that an old model cannot be applied to a new region with fewer iterations. The number of iterations was dependent on how difficult it is to differentiate non-forest types from forest types. In this study, the difficulty in differentiating non-forest types from forest varied across the 4 regions, so each went through several iterations. This process underscores the importance of "localized" forest differentiator models.



Modeling

The localized forest differentiator model for each region is a deep learning model which takes imagery data and outputs a probability that a pixel is forest.

The localized forest differentiator model is a proprietary neural network developed by Enveritas. Unlike typical deep networks used for image classification, the Enveritas model is capable of handling a myriad of multispectral bands beyond regular RGB. The model introduces several innovations to the neural network architecture that augments its abilities to handle satellite imagery. These include: a custom tensorflow module to drop out an entire multispectral source instead of just single data bands during model training, satellite source normalization to preserve relative band value differences while achieving effects of normalization, and embedded recurrent neural network for radar time series (gated recurrent units[5] within the recurrent module).

Figure 5 shows results from an experiment which demonstrate the advantages of the Enveritas proprietary neural network architecture.

The deforestation model is an ensemble model that takes in the same inputs as the localized forest differentiator model except it takes in inputs for multiple years and outputs a probability that a pixel has been deforested in 2019.

Validation

We use a combination of two approaches to assess the Enveritas approach against GFW:

- Visual interpretation (VI)
- Ground truthing

[5] Cho, Kyunghyun, et al. "On the properties of neural machine translation: Encoder-decoder approaches." *arXiv preprint arXiv:1409.1259* (2014).

Geographical sampling was conducted by dividing Enveritas 2019 deforestation events map and GFW 2019 tree cover loss events map into 30m by 30m tiles and randomly drawing tiles using simple random sampling.

We assess different types of pins:

- A land cover pin is a pin that GFW considers most likely forest in 2020. They are drawn randomly from pixels that >75% GFW tree cover in 2000 and still not deforested over 20 years. The purpose is to compare GFW and Enveritas' abilities to differentiate forest and plantation. For Côte d'Ivoire, we use 50% tree cover as the threshold as only 1% of land passes the 75% threshold.
- A forest pin is a pin that Enveritas or GFW predicts to be forest. The pin is generated by dropping it in large areas where Enveritas detects to be forest but not GFW or vice versa. The purpose is to verify where true forests are. This is only used in Côte d'Ivoire, where the map of forest varies widely between sources, to verify where true forest is.

Ground truth survey and protocol

The ground truth survey is the same for all pins. At a pin, enumerators record land types, estimate age, measure canopy cover and take photos.

If an enumerator is unable to reach the pin, they will complete the survey at their current location and report whether they can "observe" the pin location. They'd proceed to find a nearby pixel with similar land type as the pin to conduct a second survey.

Answers were collected on a smartphone.

RESULTS & DISCUSSIONS

In total, 2100 pins were visually inspected with 715 ground truth surveys conducted from 29 October 2020 - 26 January 2021 (Table 2, Table 3, Figure 9).

The Enveritas deforestation rates for Dak Nong, Côte d'Ivoire, Ethiopia coffee regions, and Aceh and North Sumatra are 0.26%, 0.23%, 0.11%, 0.36% respectively, while the GFW tree cover loss rates are much higher at 0.43%, 1.17%, 0.21%, and 0.67% (Table 5). However, within forest areas, GFW reports 12-32% less deforestation compared to Enveritas. Forest and deforestation maps are depicted in Figure 10 and Figure 11. Refer to Table 4 for forest size data.

For differentiating forests from plantations, model precision and recall on land cover pins are shown in Table 6. The Enveritas model achieves ground truthed precision of 80-94% while the GFW achieves precision of 45-75% in differentiating forest from plantation.

Finally, across all 4 regions, the average Enveritas detected deforestation cluster size is statistically smaller at the 99% confidence level than that detected by GFW. In Ethiopia, where small encroachment is the paradigm of deforestation, the average Enveritas deforestation cluster size is less than a fifth of the average GFW tree cover loss cluster size.

GFW overestimates deforestation by inaccurately classifying plantation agriculture as forest

Most of the difference is due to GFW's inability to differentiate plantations from forest (Figure 12, Table 7). In Aceh and North Sumatra, 55% of GFW's 2019 tree cover loss events are due to palm, rubber or other plantations being removed. In Dak Nong, it is 27% (coffee and rubber), Côte d'Ivoire it is 50% (cocoa, coffee, rubber, palm), and Ethiopia it is 36% (coffee, banana). Furthermore, even at high tree cover % cutoffs, GFW is incapable of differentiating plantation from forest (see appendix).

The GFW approach is to identify "tree loss" rather than "deforestation". By including plantations, GFW's data is incomparable between regions since regions differ by crop types and agronomy practices. For instance, among the 4 regions studied, Aceh and North Sumatra had the highest amount of GFW overestimation due to its large palm and rubber plantations being commonly mistaken as forest. In contrast, Dak Nong has the least amount of GFW overestimation because they do not have too many crops that register high tree cover besides coffee intercropped with pepper stands. While GFW makes clear their estimates refer to "tree loss" rather than "deforestation", many civil society organizations that then use the GFW information are unaware of this nuance, and interpreting the data as "deforestation" is common in public discussion.

Going forward, GFW will continue to overestimate deforestation since their 2000 tree cover % data cannot distinguish forest from plantations. At land cover pins, areas where GFW still considers to be forest in 2020[6], many are already non-forest - 25% in Aceh and North Sumatra, 45% in Ethiopia coffee-growing regions, 55% in Côte d'Ivoire and 27% in Dak Nong (Figure 13).

There are two main reasons for GFW's inability to remove plantations. Firstly, plantations register high tree cover % like forests. This characteristic manifests itself in different ways across the region. In Dak Nong, coffee is commonly intercropped with pepper and the combination registers high GFW tree cover % (Figure 14). In Côte d'Ivoire, GFW confuses old-growth cocoa and rubber with forest (Figure 15). This is evident in that the average age of landcover pins with cocoa and rubber is 16-21 years old, suggesting that GFW couldn't differentiate plantation from forest in their 2000 base tree cover map. The situation is similar in Sumatra, where the average

[6] >75% tree cover, no tree cover loss event since 2000

ages of palm and rubber are 16 years old (Figure 16). In Ethiopia, the practice of intercropping coffee and banana underneath forest shade trees results in high tree cover % (Figure 17). In contrast to previous regions, the average reported age of non-forest land cover pins is 4-15 years old, suggesting the replacement of forest by crop is recent and missed by GFW. Refer to Figure 18 for age data of the different land types.

The second reason is that GFW is unable to discriminate vegetation height. This problem is particularly evident in Côte d'Ivoire since the country has savannah and shrub biomes which can look like forest but are not tall or dense enough to be classified as forest (Figure 19). In Côte d'Ivoire, 26% of 2019 GFW tree cover loss events are due to shrub removal.

The Enveritas approach overcomes these issues by building a localized forest differentiator for each region. The Enveritas localized forest differentiator model achieves better precision than GFW (Table 6). We verify our model's ability to discriminate vegetation height in Côte d'Ivoire and show we can capture forests across Côte d'Ivoire, whereas GFW only finds forest in the south and west[7] and is unable to capture northern and eastern savannah regions (Figure 20). GFW's underestimation of forest in savannah regions is confirmed by FAO. Figure 21 shows a savannah area where GFW tree cover is incapable of capturing forest because it cannot differentiate vegetation height.

These results suggest that localized forest differentiator models are an integral part of deforestation monitoring.

Enveritas detects more deforestation than GFW in forest areas
When we overlap our forest map with GFW tree cover loss pins, we found more deforestation in forest compared to GFW (26% more in Dak Nong, 15% more in Côte d'Ivoire, 21% more in Ethiopia and 45% more in Aceh & North Sumatra).

Enveritas detects more small-scale deforestation than GFW
This is evident in that our average deforestation cluster size is smaller across all 4 regions, even when restricting the comparison to within forest areas (Table 8).

GFW's 30m resolution is too large to detect small-scale deforestation and thus oftentimes result in late detection. In Dak Nong for instance, intercropping and crop rotation happen frequently. Because the crop life cycles are different, farmers will remove one crop while still keeping other crops. Thus, the plantation never becomes bare land and GFW doesn't register tree cover loss. This is proven in that the average reported non-forest land cover pin is 5-7 years old despite the expectation that the average age should be greater than 20 (according to the definition of land

[7] Forest for GFW is defined as no tree cover loss event since 2000 and having 2000 tree cover >50%, which spans around 4.6 million hectares.

cover pin). Similarly, small parts of the 30m x 30m pixel are also removed in Ethiopia, where the average Enveritas deforestation cluster size is found to be less than a fifth of the average GFW tree cover loss cluster size. In Ethiopia, deforestation happens through gradual removal of small pockets of forest for cultivation and housing. Figure 22 shows an example of this pattern of encroachment where Enveritas was able to capture small-scale encroachment but GFW was unable to. In both cases, because GFW requires the entire pixel to become bare land to register a tree cover loss, it will miss these subtle changes.

GFW cannot capture reforestation

Continuous monitoring of reforested areas for deforestation is important in places where natural vegetation grows quickly (like bamboo in Dak Nong) or slowly but commonly (like the pervasive fallow land and shrub in Côte d'Ivoire) as well as in areas with active reforestation programs. In Côte d'Ivoire, for instance, the Côte d'Ivoire government has set a goal to restore forest cover to 20% of the land area by 2030.[8] But by design, GFW stops monitoring a pixel once it has lost its tree cover.

Learnings from ground truthing

Ground truthing provides feedback to improve our forest model and validates our models against others. This is particularly important in regions like Côte d'Ivoire where there is large disagreement on forest estimates (Table 9). Ground truthing in Côte d'Ivoire showed that the Enveritas forest model is more accurate than GFW because GFW confuses shrubland (low and dense woody vegetation) with tall forest and omits savannah forests. GFW's inability to differentiate shrub, savannah and forest is also highlighted in FAO's "2020 Global Forest Resources Assessment"[9]. Further details on the nature of deforestation in Côte d'Ivoire are provided in the appendix.

Ground truthing also helps our understanding of deforestation as it happens (Figure 23) and gives us insights into local deforestation methods, causes and solutions.

[8] National REDD+ Strategy (2017), Policy for Forest Preservation, Rehabilitation and Extension (2018) and 2019 Forest Code

[9] http://www.fao.org/3/ca9825en/CA9825EN.pdf

Methods of deforestation

Methods of deforestation vary across regions. In Dak Nong, deforestation happens through slash-and-burn. Deforestation in Côte d'Ivoire happens through 1) cutting trees under forest canopy to make room for crops or 2) burning trees until they die (Figure 24). Like Côte d'Ivoire, deforestation in Ethiopia happens by cutting trees under forest canopy to make room for crops like coffee and banana. Alternative methods in Ethiopia include burning trees at their root, lighting dry wood that has been placed to circle the roots, and de-barking a tree at its base (Figure 25). In Aceh and North Sumatra, deforestation happens through large-scale burning or cutting (Figure 26).

Causes of deforestation

In all 4 regions, clearing for crops is a major cause of deforestation (Figure 27). In Côte d'Ivoire, clearing is done in a progressive manner, by first starting with subsistence crops (rice, yam, cotton, banana, palm oil, or cassava) before intercropping cash crops.

In Dak Nong, other causes include logging for timber and agroforestry projects for timber, in which large scale areas (tens of thousands of hectares) are allocated to silviculture and forestry companies which exploit poor natural forests to plant acacia and eucalyptus plantations.

In Côte d'Ivoire and Ethiopia, felling trees for firewood and charcoal production is common since they are main sources of household energy. Such activities are so commonplace that they also occur in protected areas. Figure 28 shows steps in charcoal production.

In Côte d'Ivoire, another cause is bushfire for hunting agoutis and small animals sold for bush meat (Figure 29).

CONCLUSION

GFW is the only global deforestation dataset and is frequently used in civic, corporate, and government debates related to deforestation and climate change.

The intent of this study was (1) to introduce and validate the Enveritas method of deforestation accounting which combines deep learning, satellite imagery and ground truthing methods and (2) to provide the largest independent ground truthed evaluation of GFW.

Our results indicate that GFW overestimates deforestation in all 4 regions by 65 to 409%. The main cause of GFW's overestimation is its inability to remove plantations from deforestation accounting. In differentiating forest from plantations, the GFW tree cover map only achieved ground truthed precision of 45-75% whereas the Enveritas model achieves precision of 80-94%. Even by increasing their tree cover % cutoffs to 75%, there is not much improvement in GFW's ability to filter plantations.

This project showed GFW's high variation of deforestation overestimation makes it unsuitable for tracking deforestation across time and region. Instead, region-specific models coupled with ground truthing need to be developed to build a more accurate depiction of deforestation.

The Enveritas model offers such a solution. This project shows Enveritas can differentiate forest from plantation, detect smaller scale deforestation and more deforestation in forest areas compared to GFW, and provide continuous monitoring of reforested areas.

ACKNOWLEDGEMENTS

We greatly appreciate the Enveritas operations teams in Indonesia (Senthil Nathan, Ade Amanulah, Aklan Huda), Dak Nong (Hue Tran), Côte d'Ivoire (Céline Jackson, Maxime Deroubaix, and Michael Kra), Ethiopia (Morgan Minogue, Abiyi Yohannes and Yohanan Bedasso) who contributed invaluable knowledge and input in the development and implementation of the data collection methods. We owe special thanks to the team of enumerators who contributed their time and energy to data collection efforts. We also appreciate Carl Cervone, David Browning, and Maria Burkiewicz for the discussions on research design. We also thank the wider team at Enveritas for providing input or directly helping improve the data collection methods in this study.

TABLES

Table 1. Data sources used to create label data for the localized forest differentiator model

Data Source	Resolution	Year	Regions used in
Thanh et al	10m	2017	Dak Nong
Duong et al	10m	2017	Dak Nong
GLCF: Landsat Tree Cover Continuous Fields	30m	2010	Côte d'Ivoire
MightyEarth Cocoa Accountability Map	20m	2018	Côte d'Ivoire
Cocoa Living Farms map	N/A	2019	Côte d'Ivoire
Enveritas groundtruthed cocoa survey locations	N/A	2020	Côte d'Ivoire
GFSAD30: Global Food Security Analysis-Support Data at 30 Meters	30m	2015	Côte d'Ivoire, Sumatra
Global PALSAR-2/PALSAR Forest/Non-Forest Map	25m	2017	Côte d'Ivoire, Sumatra
Enveritas rubber polygons drawn by VI	N/A	2020	Côte d'Ivoire, Sumatra
High Resolution Settlement Layer	30m	2019	Côte d'Ivoire, Ethiopia, Sumatra
Enveritas groundtruthed coffee survey locations	N/A	2018-2020	Côte d'Ivoire, Ethiopia, Sumatra
Ethiopia Land Cover 2008 Scheme II	30m	2008	Ethiopia
ESA CCI Land Cover - Sentinel-2A Prototype Land Cover 20 Meter Map Of Africa 2016	20m	2016	Ethiopia
Khetami et al	30m	2017	Ethiopia
Tree plantations	N/A	2015	Sumatra
Intact Forest Landscapes	N/A	2016	Sumatra
Indonesia land cover	N/A	2018	Sumatra

Table 2: Operations summary by region.

	Dak Nong	Côte d'Ivoire	Ethiopia coffee regions	Aceh and North Sumatra
Project duration	10 - 21 Dec 2020	4 - 16 Jan 2021	13 - 26 Jan 2021	29 Oct - 15 Nov 2020 1 - 15 Dec 2020
# of enumerators	5	8	9	7
Estimated distance traveled (km)	4,777	12,000	11,990	16,660
Median distance from enumerator to target pin (m)	7.8	48.9	4.7	46.5

Table 3: Number of surveys by region and type.

	Dak Nong		Côte d'Ivoire			Ethiopia coffee regions		Aceh and North Sumatra	
	Land cover	Paired	Land cover	Paired	Forest	Land cover	Paired	Land cover	Paired
# of completed surveys	99	99	66	69	43	67	80	113	79
# of accepted surveys	97	91	54	51	42	66	78	89	44
# of accepted surveys with respondents	15	3	40	36	21	36	40	9	1

Table 4: Region and detected forest size.

	Dak Nong	Côte d'Ivoire	Ethiopia coffee regions	Aceh and North Sumatra
Total Region Area (ha)	685,544	32,246,300	3,695,292	12,254,065
Enveritas Forest Area, 2018 (ha)	215,994	4,411,911	303,958	6,380,367

Table 5: Deforestation size and rate, within region and detected forest area.

	Dak Nong		Côte d'Ivoire		Ethiopia coffee regions		Aceh and North Sumatra	
	Enveritas deforestation	GFW tree cover loss	Enveritas deforestation	GFW tree cover loss	Enveritas deforestation	GFW tree cover loss	Enveritas deforestation	GFW tree cover loss
Area within Enveritas forest (ha)	1,785	1,421	72,606	63,665	4,099	3,411	43,677	29,876
% of Forest Area	0.83%	0.66%	1.65%	1.44%	1.35%	1.12%	0.68%	0.47%
Area within entire region (ha)	1,785	2,930	72,606	376,892	4,099	7,668	43,677	81,597
% of Total Area	0.26%	0.43%	0.23%	1.17%	0.11%	0.21%	0.36%	0.67%

Table 6: GFW and Enveritas precision and recall on land cover pins.
Recall for GFW is 100% since land cover pins are sampled from GFW areas with > 75% tree cover and no tree cover loss event since 2000.

	Dak Nong	Côte d'Ivoire	Ethiopia coffee regions	Aceh and North Sumatra
GFW precision on landcover pins (%)	73	45	55	75
Enveritas precision on landcover pins (%)	94	90	80	92
Enveritas recall on landcover pins (%)	95	88	100	98

Table 7: GFW errors minus Enveritas errors, by error type and region in hectares.
Errors are calculated from the Breakdown of Enveritas 2019 deforestation events and GFW 2019 tree cover loss events figure.

	Dak Nong	Côte d'Ivoire	Ethiopia coffee regions	Aceh and North Sumatra
Overstating forest deforestation by including plantation	678	183,589	2,663	42,625
Late detection	657	13,123	424	869
Ignoring deforestation of reforested areas	276	965	32	1,395

Table 8: Average GFW tree cover loss cluster size, GFW tree cover loss cluster size within Enveritas forest, and Enveritas deforestation cluster size (ha) by region.
A cluster is a connected component - a cluster of pixels where the pixels are neighbors with each other but not neighbors with other pixels. The average Enveritas detected deforestation cluster size is statistically smaller at the 99% confidence level than that detected by GFW across all regions.

	Dak Nong	Côte d'Ivoire	Ethiopia coffee regions	Aceh and North Sumatra
GFW average tree cover loss cluster size	0.42	0.56	0.53	0.95
GFW average tree cover loss cluster size, overlapped with Enveritas forest cover	0.25	0.31	0.27	0.68
Enveritas average deforestation cluster size	0.16	0.29	0.10	0.55

Table 9. Comparison of the end 2018 forest area and 2019 deforested area across Enveritas, GFW, and FAO for Côte d'Ivoire.

	Enveritas deforestation	GFW tree cover loss at >50% tree cover	FAO[28]
Forest Area, end 2018 (ha)	4,411,911	4,746,114	2,949,900
Deforested Area, 2019 (ha)	72,606	110,394	112,900
% of Forest Area	1.65%	2.33%	3.83%

Table 10: Deforestation rate and average cluster size across Côte d'Ivoire's three forest types.

	Total area (ha)	Deforested area (ha)	Deforestation rate (%)	Average cluster size (ha)
National Parks and Reserves	726,427	1,017	0.14	68.8
Classified Forest	724,736	10,095	1.39	16.5
Other Forest	2,960,748	61,494	2.08	10.9
Overall	4,411,911	72,606	1.65	14.0

Classified Forests contain settlements and crops to varying degrees with patches of remaining old growth forest. We ground truthed 26 pins within classified forests - 13 pins are forests up to 50 years old, 2 pins are undergoing reforestation (they were previously crop, but were taken back by the forest guards to regrow), and 11 pins are crop.

We also analyze deforestation rates on the 100m buffer zone of forests and within the buffer zone (Figure 31), and find that deforestation is more than twice as likely to happen on the 100m buffer zone as within it (Table 11). In particular, Other Forests are the most fragmented as their average cluster size is the smallest. Consequently, they have the highest boundary area and are most prone to deforestation.

Table 11: Deforestation rate on 100m buffer zone and within the 100m buffer zone across Côte d'Ivoire's three forest types.

	% area on 100m buffer zone	Deforestation rate on 100m buffer zone (%)	% area within 100m buffer zone	Deforestation rate within 100m buffer zone (%)
National Reserves	26.2	0.45	73.8	0.03
Classified Forest	82.5	1.46	17.5	0.54
Other Forest	88.9	2.17	11.1	0.99
Overall	74.7	1.91	25.3	0.35

FIGURES

Figure 1: How decision trees' strict cutoff can lead to inaccurate predictions.
(A) Satellite imagery at 1.32, 99.90 shows a palm plantation where the tree cover abruptly stops in the middle because the left side has missing imagery or multispectral band values that do not meet the decision tree cutoffs. Green corresponds to areas with >75% GFW 2000 tree cover. **(B)** Satellite imagery at 6.80, 39.71 shows stripes where data is missing. The upper half is also distinctly lower in tree cover %. Intensity of green corresponds to GFW 2000 tree cover %.

A **B**

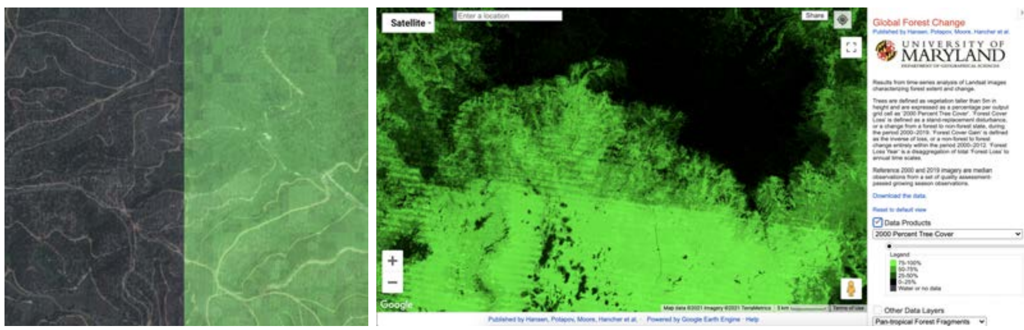


Figure 2: Higher output resolution results in sharper forest boundary. (A) Raw satellite imagery in Dak Nong, **(B)** overlaid with Enveritas 10m forest pixels as defined by the localized forest differentiator model, and **(C)** overlaid with GFW 30m forest pixels which are >75% GFW tree cover and had no tree cover loss event since 2000.

A **B** **C**

Figure 3: Map of covered areas.
(A) Dak Nong, **(B)** Côte d'Ivoire, **(C)** Ethiopia coffee regions and **(D)** Aceh and North Sumatra analyzed.

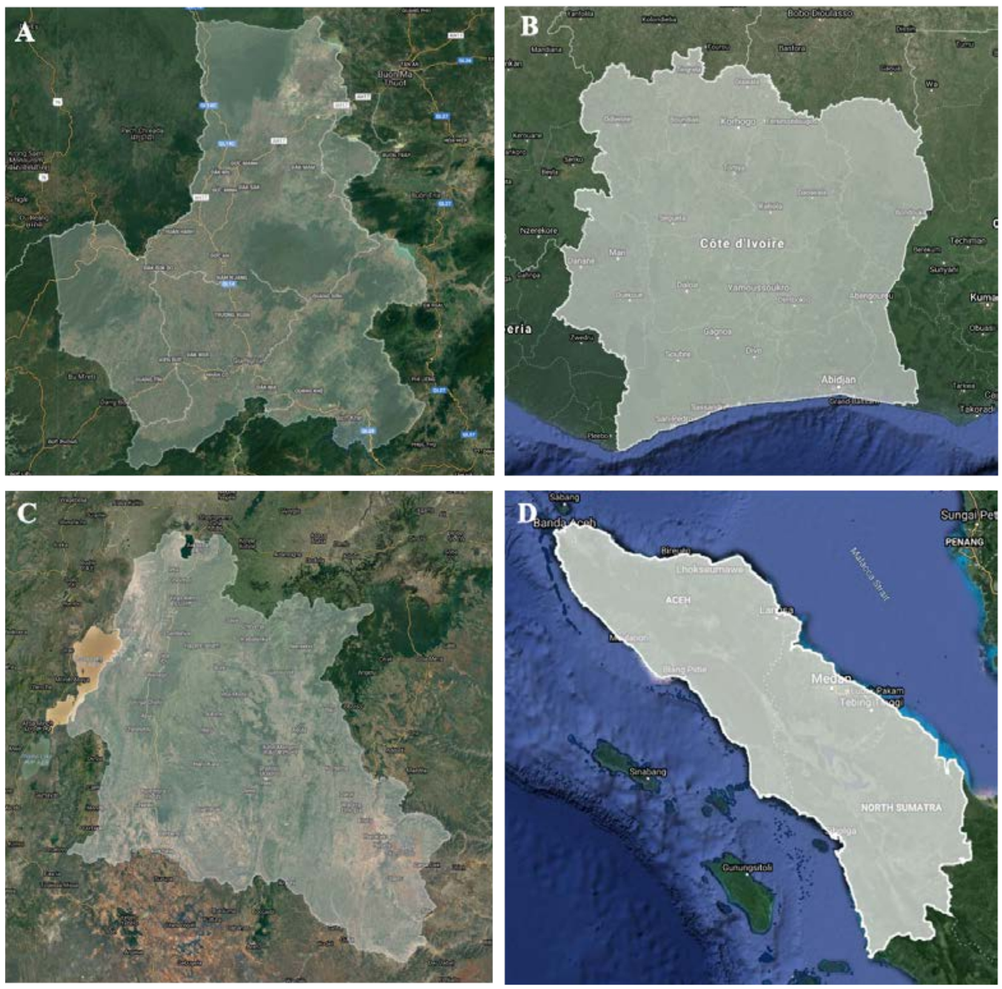


Figure 4: Enveritas forest compared with other label data sources.
(A) Satellite imagery at 5.903492, 38.731873. Green areas correspond to forest. **(B)** 10m resolution forest according to Enveritas, **(C)** 30m resolution forest according to Ethiopia Land Cover 2008 Scheme II, **(D)** 20m resolution forest according to ESA CCI Land Cover - Sentinel-2A Prototype Land Cover 20 Meter Map Of Africa 2016.

A **B** **C** **D**

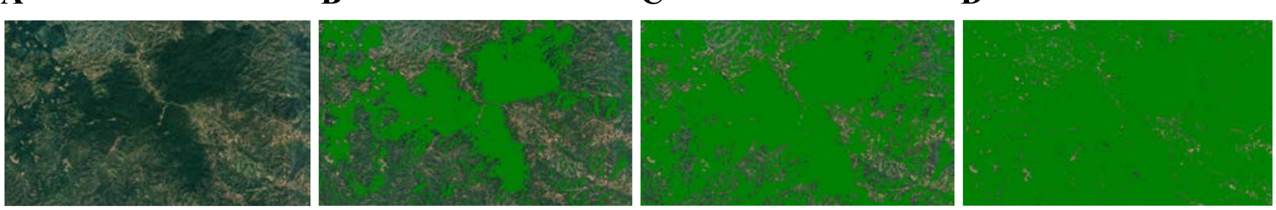
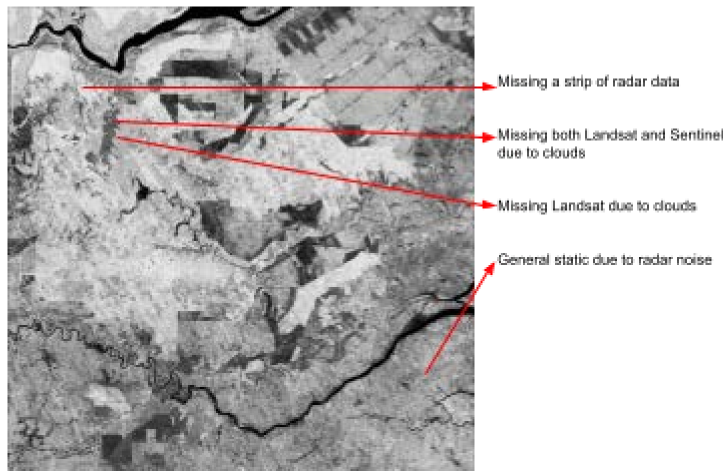


Figure 5: Enveritas proprietary neural network handles satellite imagery better than traditional neural networks.

White corresponds to high tree cover %. **(A)** Satellite imagery at -0.094421, 102.781412, **(B)** mapped tree cover % using a simple neural net whose results are very prone to satellite data availability, and **(C)** mapped tree cover % using the Enveritas proprietary network. It shows a more geographically consistent tree cover map that is more robust to the aberrations of satellite imagery input.

A

B



C

Figure 6: Visual inspection challenges in Aceh and North Sumatra.
In Aceh and North Sumatra, differentiating rubber from forest was a significant challenge. Some rubber is grown in easy to detect rows **(A, B)** but much rubber defies easy visual interpretation **(C, D)** emphasizing the need for ground truthing.

A **B** **C** **D**

Figure 7: Visual inspection limitations in Dak Nong.
In Dak Nong, differentiating rubber from forest was also a significant task. Dak Nong rubber **(A, B)** can easily be confused with forest **(C, D)** if relying only on visual inspection.

A **B** **C** **D**

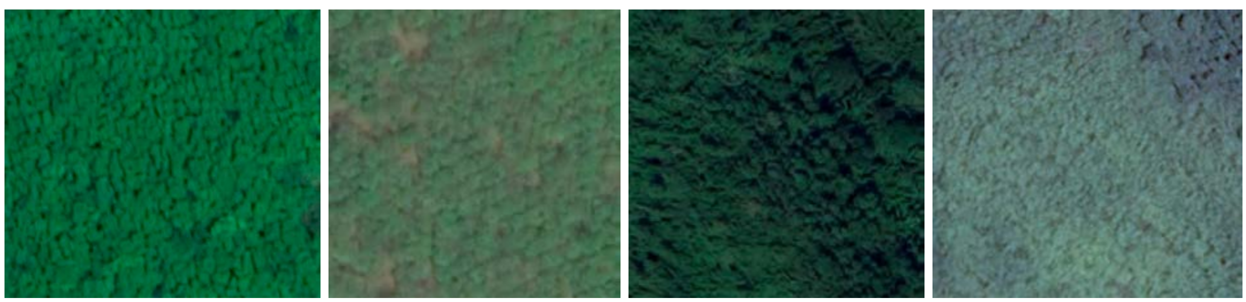


Figure 8: Visual inspection challenges in Côte d'Ivoire.
In Côte d'Ivoire, differentiating cocoa **(A, B)** from forest **(C, D)** was the biggest challenge.

A **B** **C** **D**

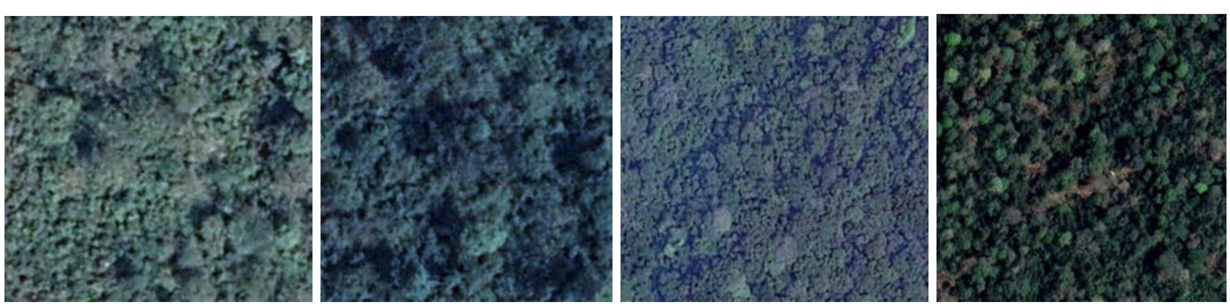


Figure 9: Map of surveyed locations.
(A) Dak Nong, **(B)** Côte d'Ivoire, **(C)** Ethiopia coffee regions and **(D)** Aceh and North Sumatra.

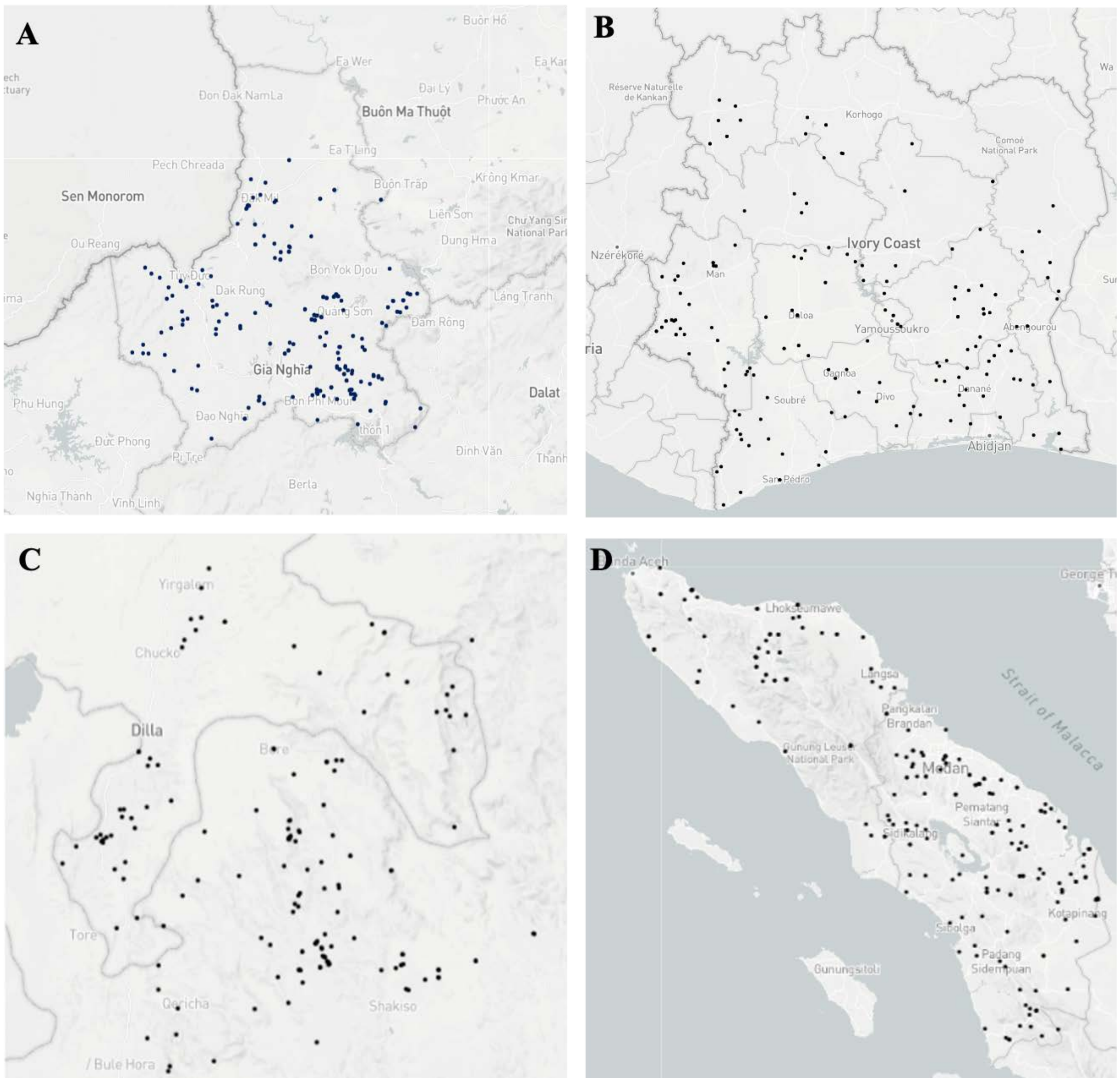


Figure 10: 2018 Enveritas forest.
Green areas are forest. **(A)** Dak Nong, **(B)** Côte d'Ivoire, **(C)** Ethiopia coffee regions and **(D)** Aceh and North Sumatra.

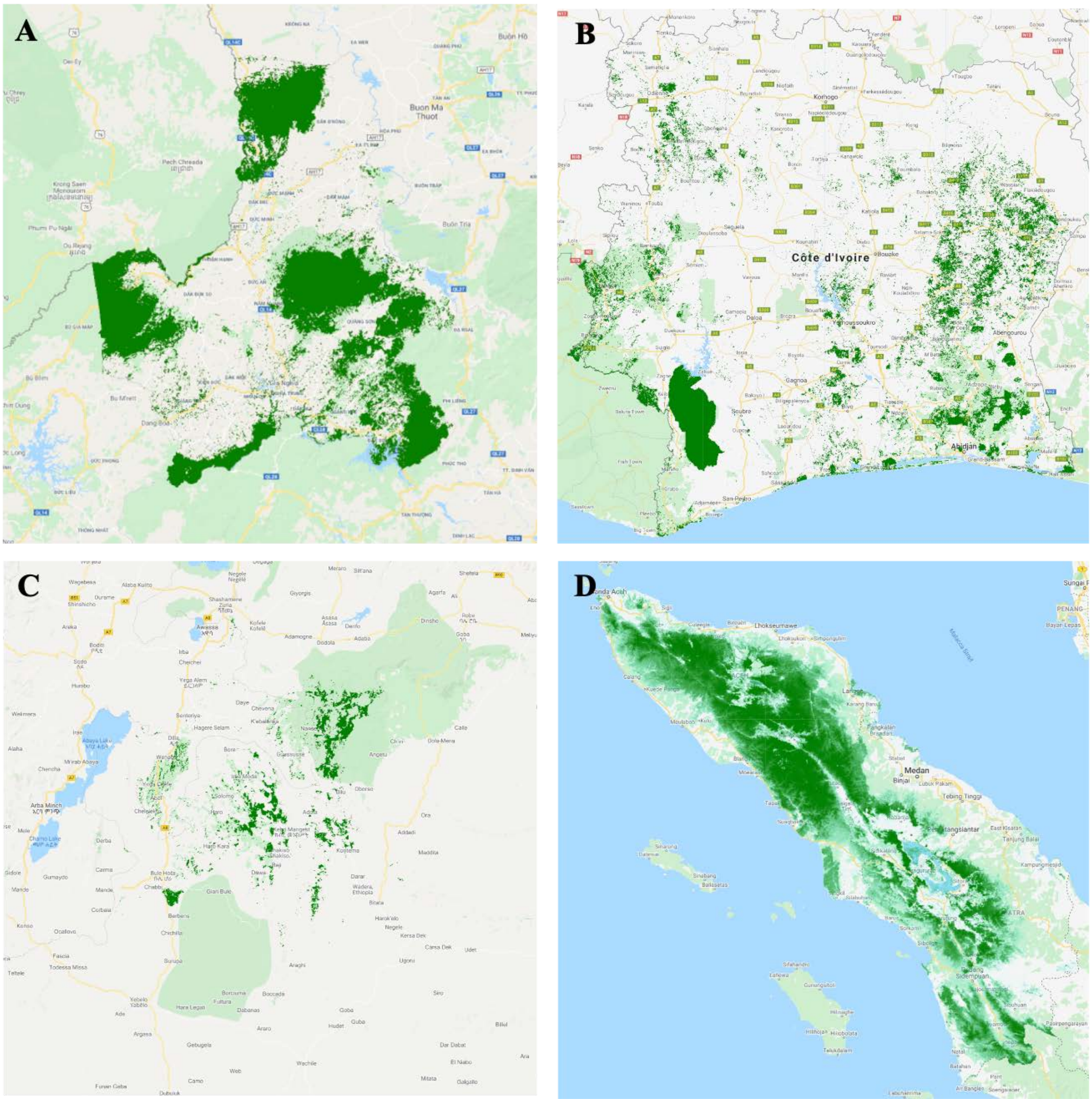


Figure 11: 2019 Enveritas deforestation map

Red areas have higher concentration of deforestation events. **(A)** Dak Nong, **(B)** Côte d'Ivoire, **(C)** Ethiopia coffee regions and **(D)** Aceh and North Sumatra.

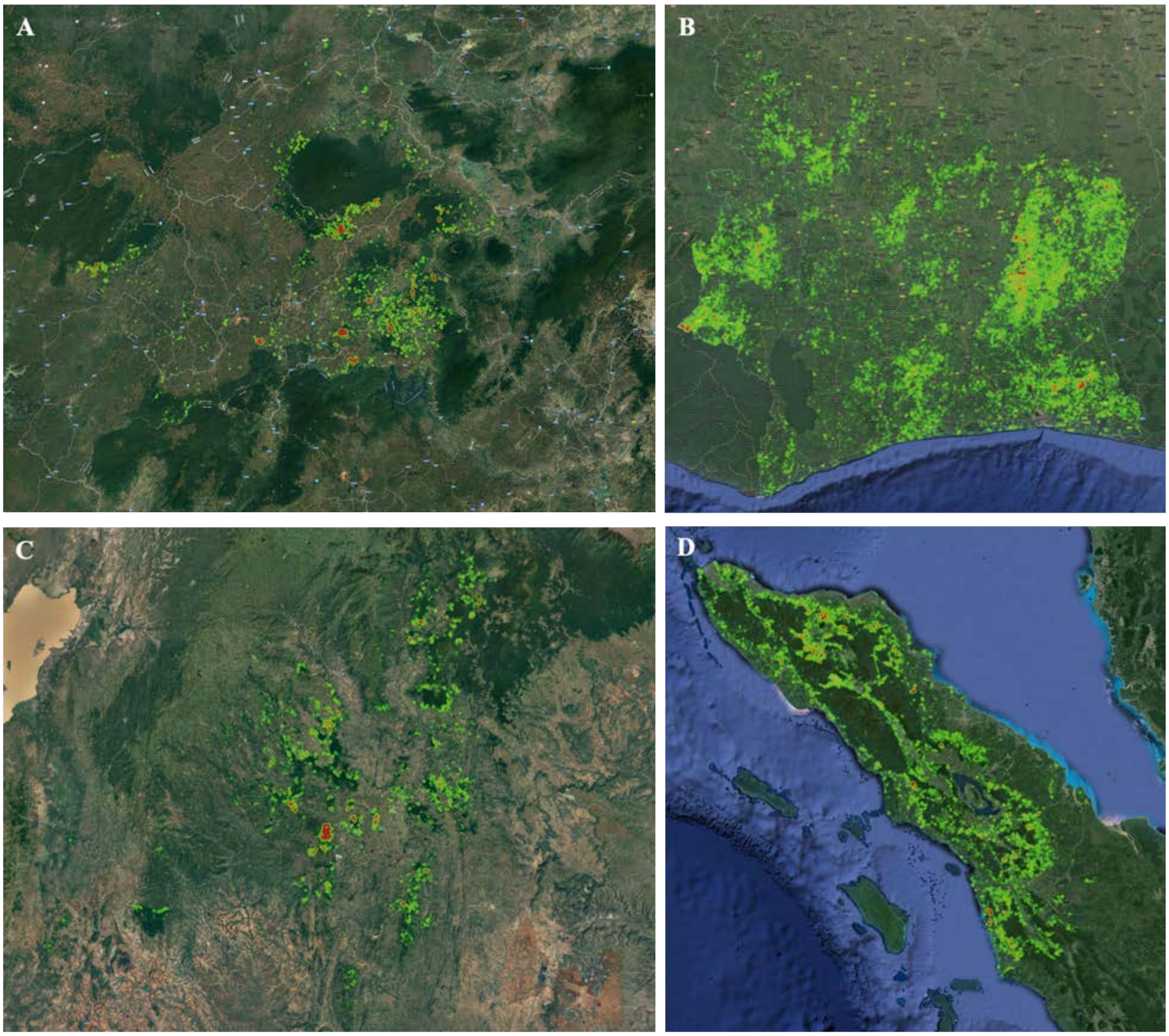


Figure 12: Breakdown of Enveritas 2019 deforestation events and GFW 2019 >10% tree cover loss events

(A) Dak Nong, **(B)** Côte d’Ivoire, **(C)** Ethiopia coffee regions and **(D)** Aceh and North Sumatra. Further details on how the figure was generated and the interpretation of the categories are provided in the appendix.

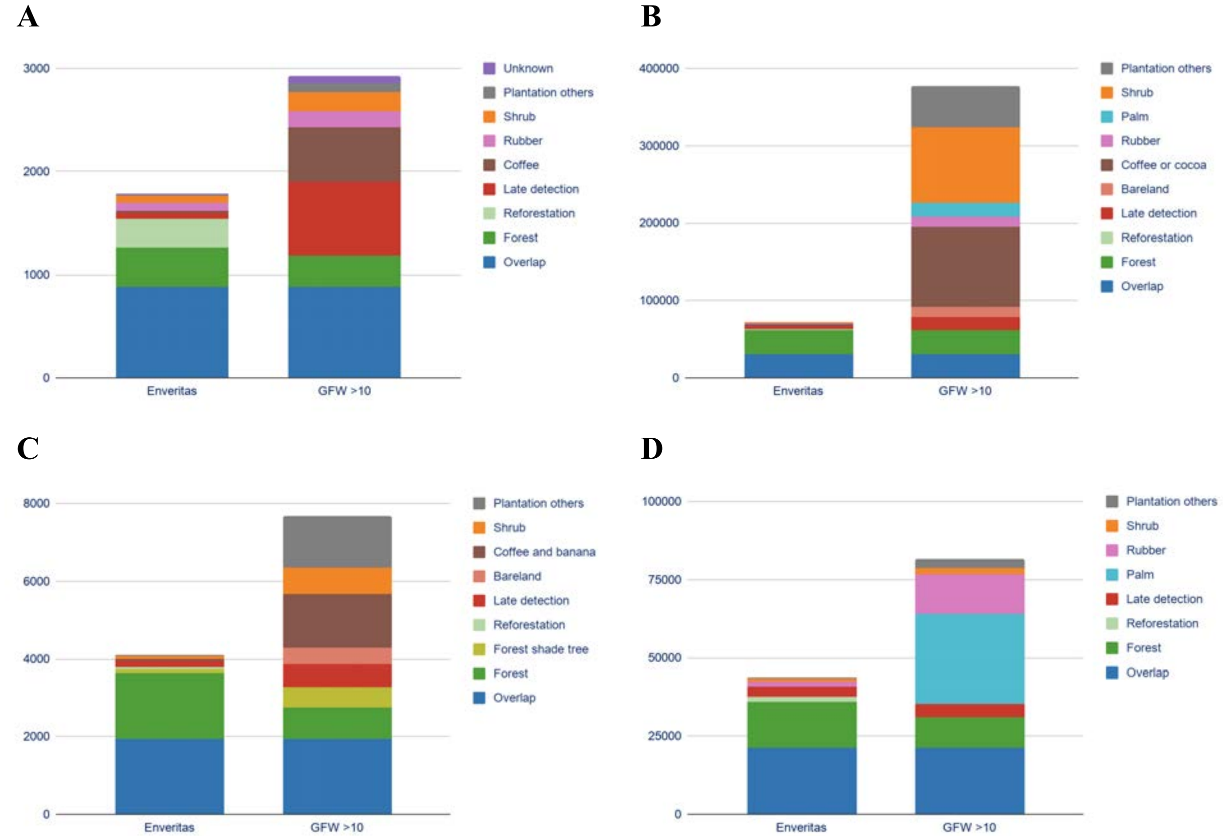
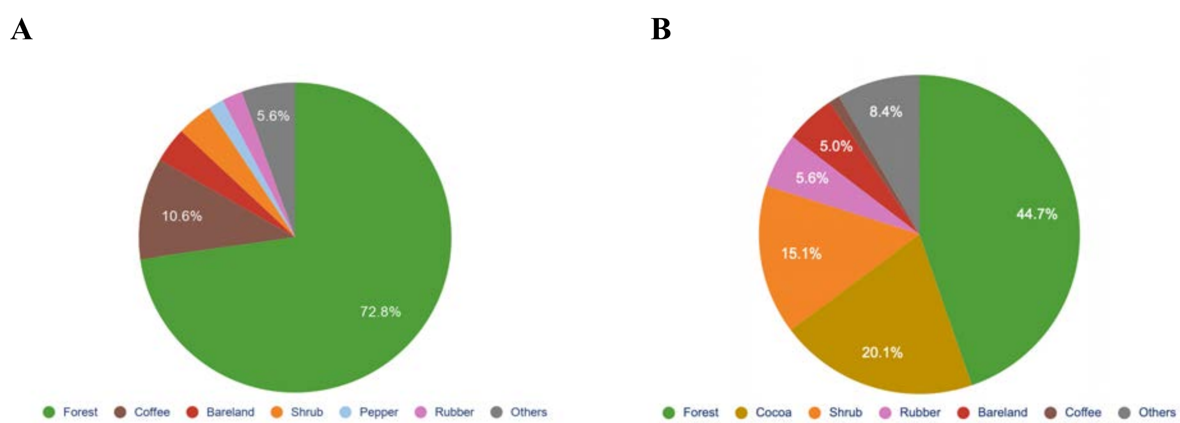
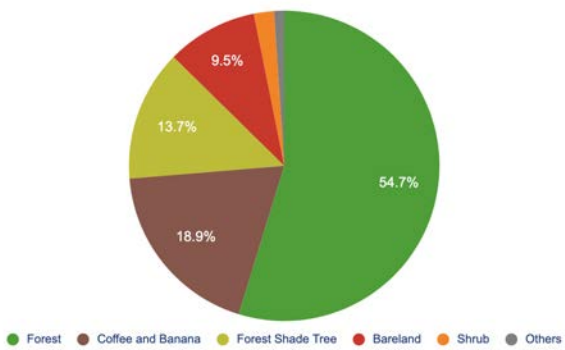


Figure 13: Breakdown of dominant land types reported at land cover pins

(A) Dak Nong, **(B)** Côte d’Ivoire, **(C)** Ethiopia coffee regions and **(D)** Aceh and North Sumatra.



C



D

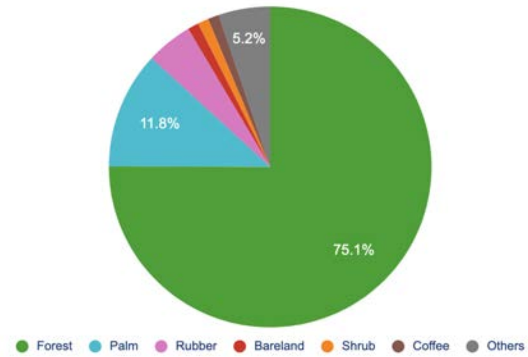


Figure 14: Coffee intercropped with pepper in Dak Nong registers high GFW tree cover %. Photos taken at a ground truthed land cover pin, where GFW reports 80% tree cover and no tree cover loss event since 2000.

East West North South Sky

Figure 15: Old growth cocoa in Côte d'Ivoire registers high GFW tree cover %. Photos of 30 year old cocoa taken near Tai National Park, where GFW reports 51% tree cover and no tree cover loss event since 2000.

Figure 16: Old palm in Aceh and North Sumatra registers high GFW tree cover %.
(A) Photos of a 25 year old palm taken at a GFW before pin, **(B)** photos of bareland taken at its corresponding GFW after pin showing the same palm from (A), **(C)** satellite imagery of (B) showing a clearing beside the palm plantation of (A).

A

B

C

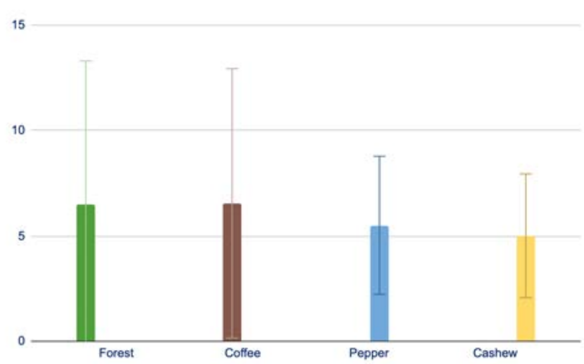
Figure 17: Coffee intercropped with banana under forest shade trees in Ethiopia registers high GFW tree cover %.
(A) Photos taken at a land cover pin of banana grown under tall forest shade trees. **(B)** Photos taken at a land cover pin of coffee grown under tall forest shade trees.

A

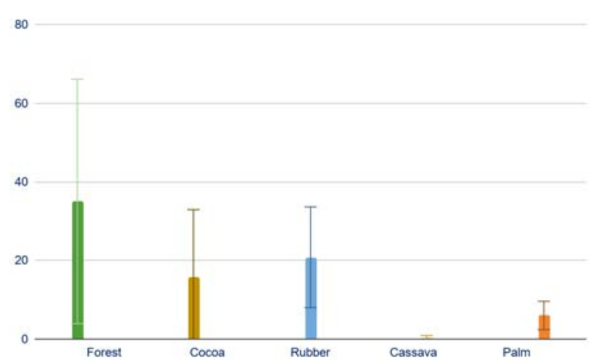
B

Figure 18: Average age of dominant land types reported at land cover pins
(A) Dak Nong, **(B)** Côte d'Ivoire, **(C)** Ethiopia coffee regions and **(D)** Aceh and North Sumatra. The survey didn't include age for shrub as it is not possible to estimate shrub age. Outliers were removed by correlating circumference, height and crown with age.

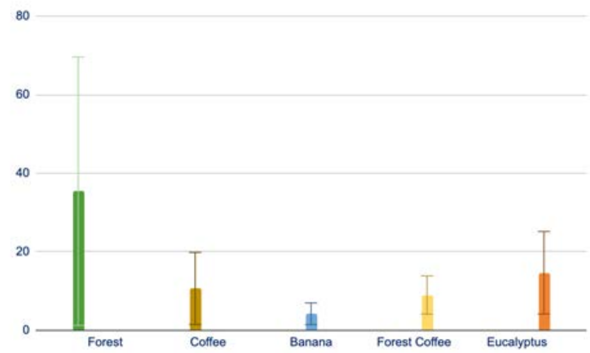
A



B



C



D

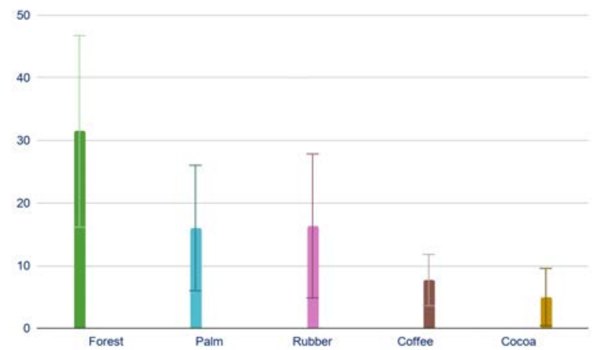
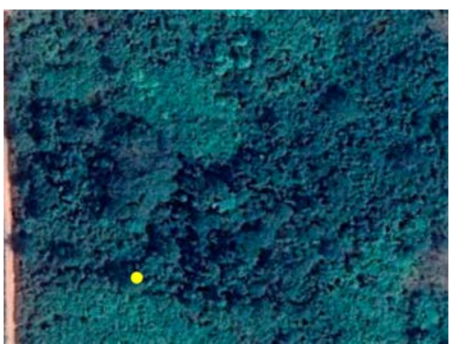


Figure 19: In Côte d'Ivoire, shrubs can look like forest in satellite imagery but are not dense or tall enough to be forest.

(A) Satellite imagery which looks like forest but **(B)** ground truthed photos show it is shrub.

A



B

Figure 20: Enveritas can discriminate vegetation height in order to detect forest whereas GFW is incapable of doing so.

(A) All ground truthed pins that are reported as forest, **(B)** subset of (A) which Enveritas also detected as forest, **(C)** subset of (A) which GFW detects as forest.

A **B** **C**

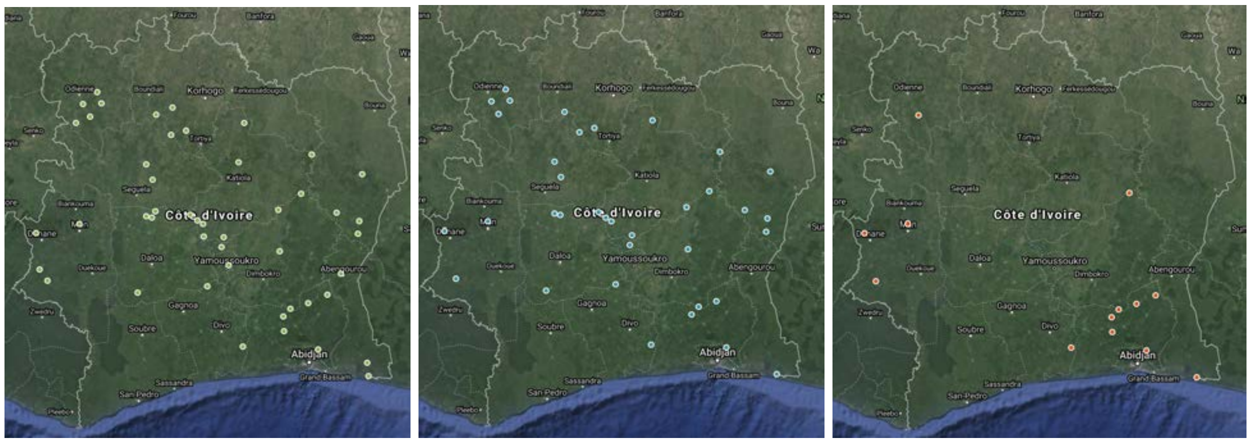


Figure 21: Comparing Enveritas forest map with forest as defined by different GFW's tree cover % cutoffs in Côte d'Ivoire's northern savannah region.

(A) Satellite imagery of northern savannah, taken at 8.691247, -4.940106. **(B)** Enveritas forest where green corresponds to forest. Red pixels are areas where there has been no tree cover loss event since 2000 and **(C)** >10% GFW tree cover or **(D)** >50% GFW tree cover.

A **B** **C** **D**

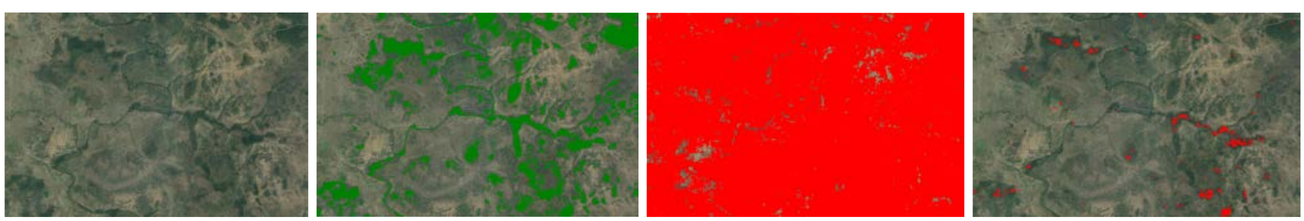


Figure 22: Example of Ethiopia small-scale deforestation where GFW cannot detect such small pockets of encroachment.

(A) Photos at an Enveritas-detected deforestation GPS showing felled logs amidst false banana.
(B) Satellite imagery showing the pin being converted from forest in December 2017 to bare land by January 2020. **(C)** GFW is unable to detect the small-scale encroachment.

A

B

C

Figure 23: Examples of recent deforestation captured through ground truthing in Dak Nong.
(A) Photos of young coffee intercropped with cassava. **(B)** Photos taken by an enumerator near (A) showing conversion to coffee. **(C)** Photos taken by an enumerator near (A) showing how common deforestation is in the area. **(D)** Photos taken at a before pin show signs of recent deforestation. The enumerator reports the deforestation event happened just a few days before the survey was conducted since the felled trees still had green leaves. Even though before pins were checked prior to ground truthing to not be bare land in 2020 imagery, many were deforested by the time we ground truthed them.

A

B

C

D

Figure 24: Deforestation in Côte d'Ivoire is through planting underneath forest or burning trees.

(A) Photos taken at an Enveritas-detected deforestation GPS, showing conversion to cocoa with signs of felled trees. **(B)** Photos taken at a landcover pin, showing felled trees amidst 1 year old cocoa and banana. **(C)** Photos taken at another location near an Enveritas-detected deforestation GPS showing young crops in the foreground and burned trees in the skyward photo. **(D)** Photo taken at an Enveritas before pin showing food crops among burned forest trees. **(E)** Photos of 1 year old cocoa at an Enveritas after pin inside Goin-Débé National Park. Note the burned trees in the skyward photo. **(F)** Forest cut and burned for yam. **(G)** Forest burned for cotton and cashews, crops which don't need shade trees.

A

B

C

D

E

F **G**

Figure 25: Deforestation in Ethiopia is through planting underneath forest and burning trees.

All photos were supplied by enumerators. **(A)** Photos showing banana grown under tall forest shade trees, taken at a land cover pin. **(B)** Photos showing coffee grown under tall forest shade trees, taken at a land cover pin. The corresponding satellite imagery shows how the plot can easily be mistaken as dense forest. **(C)** Photos of the practice of burning and cutting trees to expand a teff farm into natural forest. **(D)** Photos taken near an Enveritas after pin showing a teff farm expanding into natural forest, with old growth trees in the background and felled and burnt trees in the foreground.

A

B

C

D

Figure 26: Deforestation in Aceh and North Sumatra is through large-scale burning or cutting.

(A) Photos of forest recently burned where the enumerator reports seeing coffee seedlings, taken at an Enveritas after pin. **(B)** Photos of clearing of large plots as supplied by our enumerators.

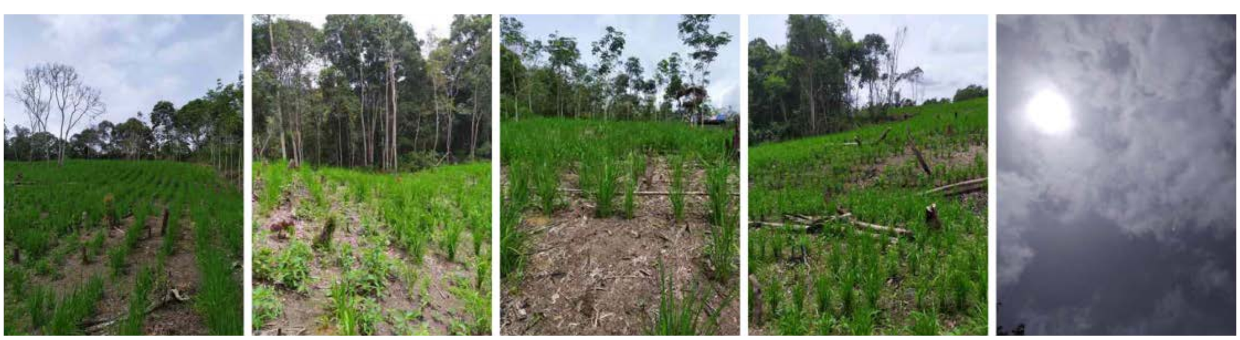
A

B

Figure 27: Deforestation due to crop.

(A) Dak Nong photos showing forest in the background, which has been recently converted into 2 year old rice paddy at an Enveritas after pin. (B) Côte d'Ivoire photos of 2 year old coffee intercropped with 8 year old cocoa and cassava taken inside a Classified Forest. There are large burned trees in the background, indicating the area was forest in the past. **(C)** Ethiopia photos of 3 year old coffee against a background of old-growth forest and a felled log. **(D)** Aceh and North Sumatra photo showing forest recently cut for 1 year old palm at an Enveritas before pin.

A



B

C

D

Figure 28: Deforestation to make charcoal.
All photos were supplied by enumerators. **(A)** Trees being burned for charcoal. **(B)** Felled trees being gathered for making charcoal. **(C)** Heaps of charcoal near Aboisso.

A

B

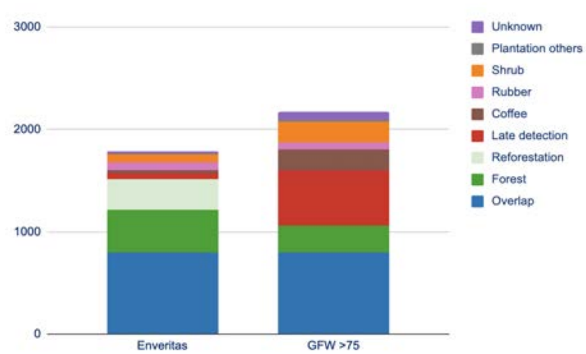
C

Figure 29: Deforestation to hunt.
Photos of bushfire aftermath in northern Côte d'Ivoire from Enveritas ground truthing.

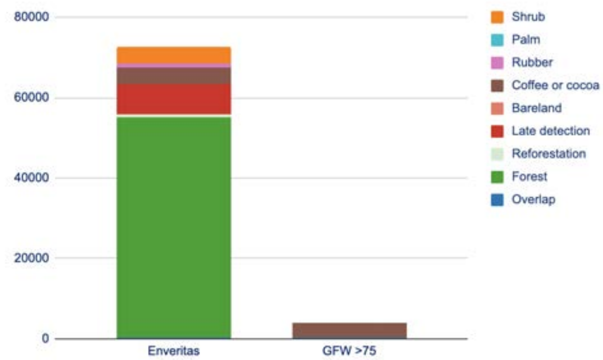
Figure 30: Breakdown of Enveritas 2019 deforestation events and GFW 2019 >75% tree cover loss events

(A) Dak Nong, **(B)** Côte d'Ivoire, **(C)** Ethiopia coffee regions and **(D)** Aceh and North Sumatra.

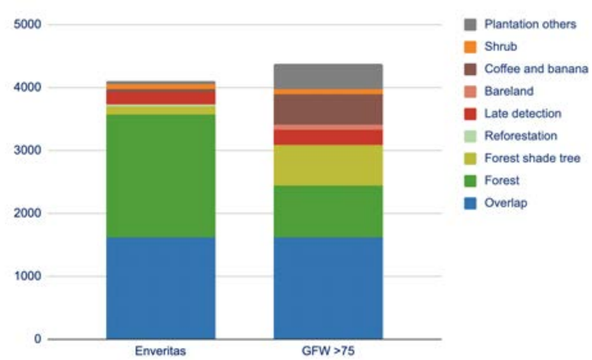
A



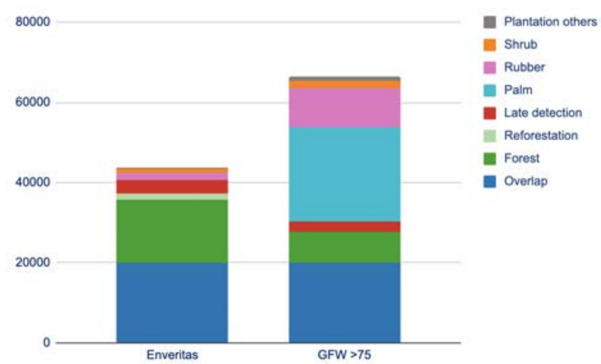
B



C



D



A counter argument can be made for Ethiopia coffee-growing regions that since the two bars are roughly the same in Figure 35 (C), GFW >75% tree cover can be a good proxy for true deforestation in Ethiopia coffee-growing regions.

This view is flawed as the accuracy of GFW>75% is contingent on the pattern of deforestation. As shown in the Enveritas bar chart in Figure 13 (C), much of the tree cover loss in 2019 is due to forest loss, but will not always be the case. Figure 14 (C) shows that much of what GFW thinks is remaining forest has already been converted to coffee (19%), forest shade trees (14%) or bare land (10%). If the deforestation pattern shifts from cutting forest to removing coffee and forest shade trees, or expanding bareland, GFW>75% will not remain a good proxy.

Figure 31: Satellite imagery on Tai National Park boundary at -6.9593, 5.4775, overlaid with black 100m buffer zone.

APPENDIX

Appendix 1. Pin shifting criteria
After the first round of ground truthing, we identified multiple operational challenges including difficult terrain, remote pins and poor infrastructure. To counter these challenges, we implemented a pin shifting criteria:

If the pin is a land cover pin, it can be shifted closer to a road as long as the new pin is within 1km of the original pin and is of the same land cover type upon inspection of high resolution satellite imagery.

-

Appendix 2. Data acceptance criteria
Some surveys were not conducted at the exact pin location because of the pins' remoteness. Thus we created the following set of criteria for survey approval. Surveys that do not satisfy any of the following are removed:

- survey distance to pin is less than 30m, survey GPS accuracy is less than 10m, and the enumerator was able to reach the pin
- survey pin is a land cover pin, its distance to pin is less than 1km, and the survey pin location also satisfies the land cover pin criteria (>75% GFW tree cover and no tree cover loss event since 2000)
- survey distance to pin is less than 500m, its reported land type matches the land type visually interpreted at the pin, and enumerator was able to reach or observe pin location

For surveys accepted and where the enumerator cannot reach but “observed” the pin, we conduct a second manual review process. We remove cases where the enumerator cannot see the pin location because of something blocking their line of sight.

Appendix 3. Figure 12 methodology and interpretation

We re-train the visual interpretation team using ground truthing results by corresponding the reported land types with satellite imagery. We visually interpret ~400 randomly sampled pins from Enveritas 2019 deforestation and GFW tree cover loss pixels for each region and categorize the pins into:

- Overlap: Pixels that both Enveritas and GFW found.
- Major land type categories such as forest, coffee, rubber, palm
- Reforestation**:** Pixels that GFW claims to have had a tree cover loss event before 2015 but Enveritas found to have another deforestation event in 2019. GFW would not be able to detect these since they assume a pixel can only be converted to bare land once.
- Late detection: Pixels that were converted to bare land in 2018 already.
- Shrub: Pixels that contain natural vegetation that is lower than the 5m required to be categorized as forest. It does not include land that is predominantly under agricultural or urban land use.
- Plantation others: Plantation pixels that do not fall cleanly into the other crop categories.
- Unknown: Pixels for which we were unable to determine whether they contain plantation or forest.

Appendix 4. Plantations are not entirely removed even at high GFW tree cover % cutoffs

Deforestation accounting methods like Quantis suggest using GFW >10% tree cover loss events as a way to measure forest tree cover loss.[10] Quantis advises that depending on the density of the canopy (e.g. Amazon tropical forest is more dense than African savannah), the tree cover % cutoff can be modified.

We suggest that in order to track deforestation, a model has to be explicitly made to discriminate plantations from forest. Approaches which aim to remove plantations by simply setting the GFW tree cover % threshold do not work. By comparing Figure 13, which shows GFW >10% tree cover loss events, and Figure 35, which shows GFW >75% tree cover loss events, we observe two findings:

- Increasing GFW tree cover % doesn’t remove errors such as late detection or confusing plantation tree cover loss with forest tree cover loss.
- The cutoff for tree cover % is hard to set, as shown in the drastic drop between GFW 2019 tree cover >10% events and GFW 2019 tree cover >75% events.

[10] https://quantis-intl.com/report/accounting-for-natural-climate-solutions-guidance

Appendix 5. Nature of deforestation in Côte d'Ivoire

Forests in Côte d'Ivoire can be classified into 3 types - forest in National Parks and Reserves, Classified Forest and Other Forest.

As shown in Table 10, most of Côte d'Ivoire forest is in Other Forest, outside of National Parks and Reserves and Classified Forest. Other Forest has the highest rate of deforestation at 2% and also the smallest cluster size, meaning that it is very fragmented. National Parks and Reserves have the lowest deforestation rate.