Establishing Best Practices and Evaluating Data and Methods for Forest Monitoring

Alliance
Bioversity & CIAT

CGIAR

World Cocoa Foundation

INFORMATION	
Organization	CIAT
CIAT contacts	Louis Reymondin (l.reymondin@cgiar.org) – Senior Scientist - Data Driven Sustainability research program leader Thibaud Vantalon (t.vantalon@cgiar.org) – Scientist Phuong Luong (p.luong@cgiar.org) – Analyst Jhon Jairo Tello (j.j.tello@cgiar.org) – Senior Analyst Jorge Perez (j.perez@cgiar.org) - Analyst
WCF contacts	Michael Matarasso (michael.matarasso@worldcocoa.org) – Impact Director and Head of North America Elizabeth Howard (elizabeth.howard@worldcocoa.org) - Environmental Manager
Version	15th of April 2025

Acknowledgment

The Alliance of Bioversity International and CIAT would like to express its sincere thanks to the World Cocoa Foundation (WCF) and the members of the WCF SBTi and EUDR Task Forces: Baronie/Cemoi, Barry Callebaut, Blommer, Cargill, Clasen Quality Chocolate, Cocoasource, Delfi Unlimited, ECOM Agrotrade, ETG / Beyondbeans Ferrero, GCB, General Mills, Halba, JB Cocoa, Lindt & Sprüngli, Mars, Mondelēz, Natra, Nestlé, OFI, Starbucks, Sucden, The Hershey Company, Touton, and Unilever for their valuable contributions to the development of this methodology.

This report reflects the efforts of a wide group of experts from the Alliance of Bioversity International and CIAT. We would especially like to thank the following contributors for their input in their respective fields of expertise: Christian Bunn (Criteria Review), Tiffany Talsma (Review).

We are grateful to all collaborators for their support and insights throughout this process.

Table of Contents

List of Figures

List of Tables

Introduction

The urgent global need for robust remote sensing solutions has grown quickly in recent years, driven by international regulations and sustainability commitments. The European Union Regulation on Deforestation-Free Products (EUDR), the forthcoming Corporate Sustainability Due Diligence Directive (CSDDD), the Corporate Sustainability Reporting Directive (CSRD), and global net-zero ambitions underscore the critical necessity for advanced remote sensing tools and high accuracy maps to assess compliance with zero deforestation value chains, land-use changes monitoring, and greenhouse gas (GHG) emissions targets.

The current landscape is marked by a variety of datasets and services to determine compliance and estimate non-compliance risks. However, significant challenges remain. First, there is no global consensus on the quality and limitations of existing datasets and systems specifically for the cocoa sector, which hinders their effectiveness in diverse landscapes. Second, there is limited transparency regarding the accuracy and performance of these systems across different contexts. These gaps strongly limit the applicability of such solutions.

To begin to address these limitations, collaboratively, the World Cocoa Foundation (WCF) with the Alliance of Bioversity International and CIAT, as well as key partners such as WCF members, international research institutes, and national institutions in Ghana and Côte d'Ivoire, have developed a comprehensive four-tiered data assessment framework. This framework was used to evaluate the quality and performance of 21 of the most referenced publicly available critical datasets related to (1) land cover and land use change (LUC), (2) carbon and biomass estimations, and (3) tree-planting initiatives (agroforestry and reforestation) together with 4 platforms that use and combine these datasets to provide tailored analysis for a specific supply chains. By assessing commonly referenced and publicly accessible datasets and platforms, as well as Satelligence, a private solution provider (included because of their deforestation risk assessment work for WCF and several members on EUDR), this project ensures that these data layers can be effectively utilized for value chain assessments and compliance monitoring.

The selected datasets and platforms were tested against four key criteria in Ghana and Côte d'Ivoire: **accuracy**, assessing how well the data represents reality on the ground; **completeness**, evaluating geographic, temporal, and categorical coverage as well as resolution suitability for cocoa landscapes; **data management**, examining metadata quality, accessibility, licensing, interoperability, and long-term usability; and **inclusivity**, reviewing efforts to incorporate smallholders and vulnerable stakeholders in dataset.

This white paper presents a description of the assessment criteria first (full list in appendix), followed by sections dedicated to each of the three categories of analysis (1) land cover and land use change, (2) carbon and biomass, and (3) tree-planting initiatives. Each section includes:

i. a detailed overview of the methodology used to assess the relevant datasets and platforms for each category,
ii. results of the analysis,
iii. a discussion of findings.

This is followed by a section on platforms and sections for conclusions and recommendations that explore the potential for these datasets and platforms to enable compliance and support sustainable land-use practices.

Assessment Criteria

The selection of datasets to be explored, the criteria for assessing their quality, and the development of the ranking system were built through a consultative approach. This collaborative process involved contributions from a diverse range of stakeholders, including WCF members, national institutions in Côte d'Ivoire and Ghana, and international organizations. Their ideas, expertise, and feedback were instrumental in refining and consolidating the quality assessment process during its definition. By incorporating insights from these key stakeholders, the methodology was tailored to address the unique challenges and priorities of different contexts, ensuring a comprehensive and relevant evaluation framework.

The assessment of datasets and platforms in this study was based on four key areas: Accuracy, Completeness, Data Management, and Inclusiveness. These criteria ensure a comprehensive evaluation of the datasets' reliability, coverage, usability, and representativeness. A ranking system was applied to assign weighted scores, prioritizing critical factors like accuracy, to determine the overall suitability of each dataset.

Accuracy

Accuracy evaluates how well the datasets capture real-world land cover and land cover changes, ensuring the reliability of classifications. This includes:

- **Overall Accuracy**: How well the dataset performs across all land cover categories.
- **Precision**: The ability of the dataset to correctly classify specific land types (e.g., cocoa, forest). It is equivalent to 1 – commission error.
- **Recall**: The ability to capture the full extent of areas of each land cover. It is equivalent to 1 – the omission error.

For datasets that could not be independently validated—such as those for carbon and biomass monitoring and tree planting—accuracy was assessed based on the figures self-reported in the associated scientific publications. These reported metrics were carefully reviewed, summarized, and included in this report. To provide a clearer interpretation of their suitability, each dataset was assigned a usability score ranging from Excellent to Very Limited.

The *usability* evaluation process of datasets that could not be independently validated, considered three main aspects. First, validation quality was examined to determine whether the dataset had been validated using appropriate methodologies and whether its reported accuracy was robust. Second, methodological limitations were assessed, identifying constraints in the dataset's creation that could affect its applicability for land cover, carbon, or tree monitoring. Finally, reported limitations were reviewed to highlight any documented issues that might impact the dataset's intended use. Where such limitations were identified, they were explicitly noted in the report to provide users with a comprehensive understanding of the dataset's strengths and weaknesses.

Completeness

Completeness examines whether the datasets and platforms provide adequate geographic coverage, temporal resolution, and data granularity. It also assesses whether they include all relevant land cover types and agroforestry systems and if historical data is sufficient to monitor trends such as deforestation and GHG emissions. Datasets with significant gaps in coverage or missing critical land types are scored lower in this category.

Data Management

Data management assesses the quality of metadata and data management practices, ensuring adherence to the FAIR principles (Findable, Accessible, Interoperable, and Reusable).

Inclusiveness

Inclusiveness measures how well the datasets represent vulnerable and smallholder communities within cocoa-growing regions. This criterion evaluates:

- The presence of strategies to ensure fair representation in data collection.
- Mechanisms that allow smallholders to view and correct their representation in the datasets. Datasets that fail to account for or involve smallholder communities were marked down in this category, as their representativeness is critical for sustainable monitoring.

Ranking System

The datasets and platforms were evaluated using a weighted ranking system that assigns points based on performance in each criterion. Accuracy was derived from the F-score, calculated using the precision and recall metrics for each category, and was weighted twice as heavily as other criteria, reflecting its critical importance. Final scores were calculated by normalizing each topic (Accuracy, Completeness, Data Management, and Inclusiveness) and summing the weighted values.

This systematic approach ensures that the datasets selected for analysis are robust, reliable, and aligned with the goals of deforestation monitoring, carbon accounting, and tree planting assessments.

The full list of criteria used in the evaluation, and their respective weight, is provided as an annex in the Excel file *quality_assessment_results* for reference.

Land Cover and Land Use

Remote sensing offers unparalleled benefits for forest and land cover monitoring, providing a comprehensive view of vast and often inaccessible areas. Its ability to deliver consistent, large-scale data allows stakeholders to monitor deforestation patterns, detect changes in forest cover, and estimate key metrics such as biomass loss over time. The multi-temporal analysis capabilities of satellite imagery enable the identification of seasonal variations and long-term trends, critical for tracking forest dynamics and implementing timely interventions. Furthermore, remote sensing is cost-effective compared to traditional ground-based methods, enabling efficient and uniform data collection across diverse regions.

However, the limitations of remote sensing should not be overlooked. Freely available satellite sensors often have limited spatial and temporal resolution, making it challenging to detect small-scale changes or dynamic processes and challenging to determine land cover and non-forest tree type. Dense forest canopies and the spectral similarity of land cover types may also obscure features such as understory vegetation or agroforestry systems, leading to misclassification or incomplete assessments. Additionally, remote sensing data is susceptible to atmospheric conditions and requires validation through on-the-ground measurements, which can be logistically demanding and expensive. For uneven terrains or mixed-use landscapes, processing errors and the "mixed pixel" issue, where a single pixel captures multiple land cover types, can further complicate data interpretation. These challenges highlight the need for careful data selection, processing, and integration with complementary methodologies.

Agroforestry systems, and cacao in particular, present unique challenges for remote sensing. Their diverse and complex nature, often involving a mix of tree species, crops, and other vegetation, complicates classification efforts. Lower spatial resolution imagery struggles to capture small-scale practices or distinguish individual trees. Moreover, optical remote sensing methods may provide limited information on the vertical structure of vegetation, a critical aspect of agroforestry systems. Addressing these limitations requires a multi-sensor, multi-scale approach, integrating ground-based observations and advanced field surveys to enhance accuracy and reliability. Despite these constraints, remote sensing remains an indispensable tool for large-scale monitoring, and its integration with complementary techniques is essential for improving the assessment of agroforestry systems and forest dynamics over time.

In recent years, the development of publicly available global datasets has significantly advanced the ability to monitor forests and land cover at large scales. Several datasets focus on tree cover, such as Dynamic World, ESA World Cover, and the widely referenced dataset by Hansen et al., 2013. These datasets provide foundational insights into forest extent and changes over time. However, their applicability to agroforestry systems, particularly cocoa cultivation, remains limited. Cocoa, being a tree crop, requires specialized datasets that go beyond generic tree cover data. To address broader deforestation concerns, other datasets have been developed to map natural forests globally. These include initiatives such as JRC EU, Tropical Moist Forest (TMF) layers, GLAD Forest layers, and efforts by WRI/SBTN focused on identifying natural lands.

Although cocoa has not been mapped at a global scale, several country-level efforts and targeted initiatives provide valuable insights. In Côte d'Ivoire and Ghana, cocoa is included as a distinct land cover class in national mapping efforts. High-resolution satellite-based maps produced by ETH Zurich for 2020 further enhance this understanding, offering detailed insights specific to these countries. The Forest Data Partnership also recently released cocoa maps based on high resolution imagery for Côte d’Ivoire and Ghana. Additionally, private sector solutions, such as those developed by Satelligence, have created maps tailored to cocoa systems. These local datasets are critical for addressing the unique challenges of cocoa related deforestation and establishment of agroforestry systems and ensuring their integration into sustainable land-use planning and deforestation prevention efforts.

Several datasets, including TMF, Hansen et al Tree Cover Loss, GLAD, and Satelligence, provide a valuable tree cover or forest baseline along with associated tree cover loss or deforestation data. Change detection methods utilized by these datasets have proven highly efficient, with recent studies demonstrating accuracy rates ranging from 80% for publicly available datasets to as high as 95% for commercial solutions. However, these assessments are primarily designed to detect whether a change has occurred from the baseline, without necessarily determining whether the change is actual forest loss. The effectiveness of these algorithms in identifying deforestation is therefore heavily dependent on the quality of the baseline data and its ability to accurately differentiate forests from other land cover types.

The analysis is conducted in two phases: first, through a quantitative, independent quality assessment to evaluate the accuracy and reliability of the datasets; and second, through a qualitative evaluation focusing on their completeness, data management practices, and efforts to ensure inclusivity and representativity.

Selected datasets

Table 1 Selected land cover datasets, F: Datasets with Forest category available, C: Dataset with cocoa category available, NF: None-Forest Natural lands category available, TC: Tree cover instead of forest.

Title	Abbreviation	F	C	NF
Satelligence	Satelligence	+	+	+
Carte Occupation du Sol Ivoirien 2020	Official-CIV	+	+	+
The national land use map for Ghana	Official-GH	+	+	+
Cocoa Probability model 2024a and Forest Persistence v0	FDaP	+	+	
SBTN Natural Lands Map 2020 v1.1	SBTN-NL	+		+
Dynamic World V1	DW	TC		+
Global forest cover 2020 V1	JRC-EU-V1	+		
Global forest cover 2020 V2	JRC-EU-V2	+		
Tropical Moist Forests product – Annual Change v1 2020	TMF	+		
ESA World Cover	ESA-WOC	TC		
Global Forest Change 10%	GFC-10%	TC		
Global Forest Change 30%	GFC-30%	TC		
Forest extent, 2020	GLAD-Forest	+		
ETH high-resolution maps of cocoa	ETH-Cocoa		+	

Global Pasture Watch	GPW		+
Global 30-meter Land Cover Change Dataset	GLC_FCS30D	+	+

Table 1 provides an overview of selected datasets used for land cover quality assessment, highlighting their availability and relevance to three key land cover categories: **Forest**, **Cocoa**, and **Other Natural Lands**. Each dataset's capability to represent specific land cover types is indicated with a plus sign.

- **Forest**: Some datasets, including Satelligence, Official National Maps (CIV and GH), FDaP, JRC-EU-V1 and V2, TMF, and GLAD-Forest, have a specific forest category, explicitly classifying areas as forest land cover. In contrast, other datasets, marked as "tree cover" in Table 1, only define forests based on a tree cover threshold, which includes any area meeting a certain tree cover percentage. This distinction is important, as tree cover datasets may encompass plantations, agroforestry systems, or other landscapes with tree presence that do not necessarily meet the definition of a forest.
- **Cocoa**: Only a few datasets, such as Satelligence, Official National Maps (CIV and GH), and ETH-Cocoa, specifically include cocoa as a land cover class.
- **Other Natural Lands**: Datasets like GPW, SBTN-NL, DW provide other natural land cover types beyond forests.

Datasets labeled "Tree Cover" offer generalized tree cover data, which may be useful but have limited specificity for cocoa or natural forest classifications.

Method Overview

Land cover definitions

This study strictly adheres to the following specific definitions when assigning a land cover classification to a given area.

Forests: The EUDR applies the forest definition from the Food and Agriculture Organization of the United Nations (FAO). In the FAO definition, forests are "*land spanning more than 0.5 hectares with trees higher than 5 metres and a canopy cover of more than 10%, or trees able to reach those thresholds*" (Article 2, Litra 4), excluding agricultural plantations "*such as fruit tree plantations, oil palm plantations, olive orchards and agroforestry systems where crops are grown under tree cover*" (Article 2, Litra 6).

Cocoa: is defined as any land where cocoa is cultivated, regardless of the growing system, age of the trees, or management practices.

Non-forest natural lands: Following IPCC classification scheme, non-forest natural lands combine the following categories ***Non-Forest Wooded Lands*** defined as scrubland which may include national parks and wilderness recreational areas, ***Savannahs*** defined as a landscapes characterized by a mix of grasses and scattered trees, with tree cover below 10%, often utilized for grazing, and ***Wetlands*** defined as *non-Forest* marshes.

Others: Any land cover that does not fall under any of the previously defined categories.

Accuracy assessment

The accuracy assessment method employed in this study was designed using a systematic approach to evaluate the reliability of land cover classification across Côte d'Ivoire and Ghana. The process began with defining the scope of the analysis. This was done based on an initial assessment of cocoa distribution across administrative areas in both Côte d'Ivoire and Ghana. The original plan was to exclude administrative areas (districts) where no cocoa production was reported in the datasets used for the study, specifically Satelligence and official national data. However, as shown in Figure 1, nearly all administrative areas contained some level of cocoa presence according to these datasets, making it challenging to establish clear exclusion criteria without introducing potential bias.

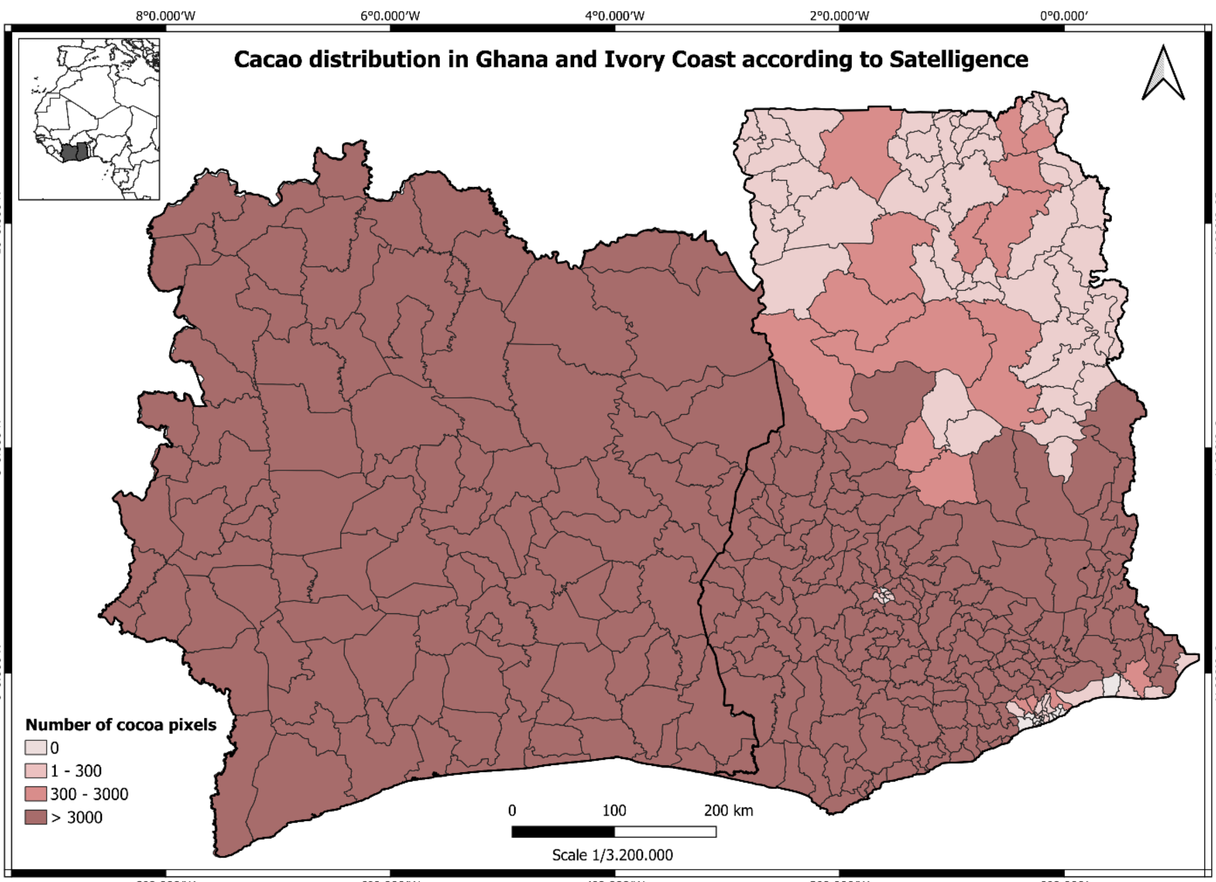

Figure 1 Cacao distribution in Ghana and Côte d'Ivoire according to Satelligence

Given these findings, we opted to conduct the analysis at the national level to ensure objectivity and consistency. Restricting the study area based on predefined thresholds could have led to subjective decisions about what should be considered noise versus real cocoa detection before conducting validation. Additionally, applying a uniform methodology across both countries was critical for ensuring that the results remained comparable.

defining the sampling area, with a focus on regions where high-resolution imagery from 2020 was freely available. Since not all areas within the study regions met this criterion, the analysis was limited to locations where such data was accessible. A grid-based sampling strategy was applied, and 500x500m cells were randomly selected. These cells were carefully reviewed by interpreters to ensure the availability of high-quality imagery (e.g., free of cloud cover and shadows) from 2020. The sampling ensured that the selected sites represented the overall class distribution, enabling a robust and unbiased assessment of land cover accuracy within the study areas.

As shown in Figure 2, the sampling methodology was based on stratified random sampling using land cover data from Satelligence. Strata were carefully defined to include key land cover classes: **Cocoa**, which was further divided into two substrata (cocoa without shade and cocoa with shade), **Forest**, which included a substratum for regrowth, **Non-forest natural lands**, and **Other** land types. This ensured that the sampling captured the full range of land cover variability in the study region. To achieve a target margin of error within ±3.5% at a 95% confidence interval for overall accuracy, sufficient sample points were selected for each stratum in accordance with the guidelines of Olofsson et al., 2014.

The validation process involved independent visual validation of each sample point using high-resolution imagery. Each point was validated independently by at least four interpreters to ensure consistency and minimize bias. In cases where disagreements occurred, a consensus process was employed, and discrepancies were discussed among the interpreters to reach a final collective decision.

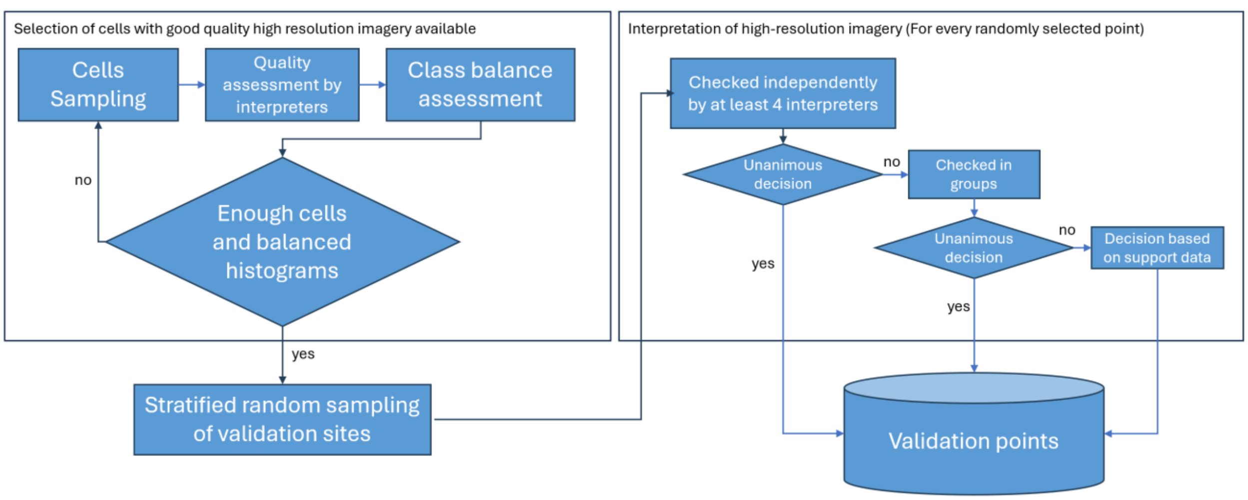

Figure 2 Overview of the sampling strategy

Accuracy metrics were calculated to provide both overall and class-specific evaluations of land cover classification. Overall accuracy was reported with a confidence interval (e.g., X% ± Y%), offering a general measure of the classification's reliability. For each land cover class, precision (1 – commission error) and recall (1 – omission error) were calculated, also with confidence intervals, to provide deeper insights into class-specific performance.

To identify the most suitable publicly available datasets, we recommend applying a two-step selection approach. First, a minimum precision threshold is set to ensure the dataset provides a reliable classification. Among datasets that meet this initial precision requirement, we then prioritize those with the highest recall to minimize the risk of omitting deforestation events. This approach balances data reliability and completeness, ensuring that selected datasets effectively capture forest changes while maintaining acceptable accuracy.

To ensure a fair and robust evaluation, we implemented measures to manage both bias and uncertainty in the assessment. Bias management was addressed by adjusting for differences between the validation strata (which were initially based on Satelligence data) and the dataset strata, ensuring that no dataset, including Satelligence, had an inherent advantage. This approach follows the methodology outlined by Stehman (2014) to maintain the integrity of comparisons across datasets. Uncertainty management was also a key consideration, particularly in estimating Satelligence's accuracy before validation, which was critical for determining the appropriate sample size needed to achieve the targeted confidence interval. To mitigate potential inaccuracies, we increased sampling by 40% for key categories, such as shaded cocoa, full-sun cocoa, forests, and non-forest natural lands. This adjustment helped reduce the impact of uncertainty and ensured a more reliable accuracy assessment.

Composite Analysis

Relying on a single land cover map yields limited results due to inherent variations in definitions, spatial and temporal resolution, and classification systems used across different datasets. Recognizing this challenge, the Forest Data Partnership (FDaP) and the AIM4Forests Programme developed and promoted the Convergence of Evidence approach as a strategy to address these inconsistencies. By integrating multiple datasets, this approach enhances accuracy and reliability, providing a more comprehensive and nuanced understanding of land cover and land use dynamics. To facilitate its implementation, WHISP, a tool hosted and developed by the FAO-led Open Foris platform, was created as a practical solution. WHISP enables users to apply the Convergence of Evidence approach effectively, ensuring more precise and reliable land cover assessments.

In this study, we tested this approach by selecting a set of datasets that individually demonstrated good accuracy for key targeted land cover classes. The primary objective of this composite map is to establish a baseline for EUDR compliance monitoring with a cutoff date set to the end of 2020. These datasets were then combined using a decision tree methodology, integrating their strengths while addressing their individual limitations. The resulting map was subsequently tested against the validation datasets to evaluate whether the quality of the combined map had improved.

In compliance-driven applications, it is necessary to assign a single, definitive land cover category to each pixel, whether it is forest, cocoa, or another category, to determine deforestation risks. While separate maps per category could provide additional confidence indicators, a final classification decision is still required for compliance monitoring. The composite approach streamlines this decision-making process while ensuring that land cover is assigned based on the most reliable available evidence.

This composite analysis is explorative and serves to test what can be achieved with such an approach rather than providing the optimal set of rules for constructing a composite dataset. A key limitation of this analysis is that the same validation data was used both to select and test datasets, potentially introducing bias and inflating accuracy estimates. Unless independent validation is conducted, there is no straightforward way to fully address this methodological limitation. However, the impact of using the same data points for both selection and evaluation is limited. This is because the actual category and location of the validation points were not used to generate the composite layer, unlike in a traditional classification model where training data directly influences the output. Therefore, while this analysis does not define the best possible composite methodology, the results still provide indicative insights into the potential effectiveness of this approach, in line with the original objectives of the assessment. We acknowledge that further research is needed to determine the optimal dataset combination for maximizing recall or precision in specific categories and to robustly validate the results.

Results

Accuracy assessment

Sampling

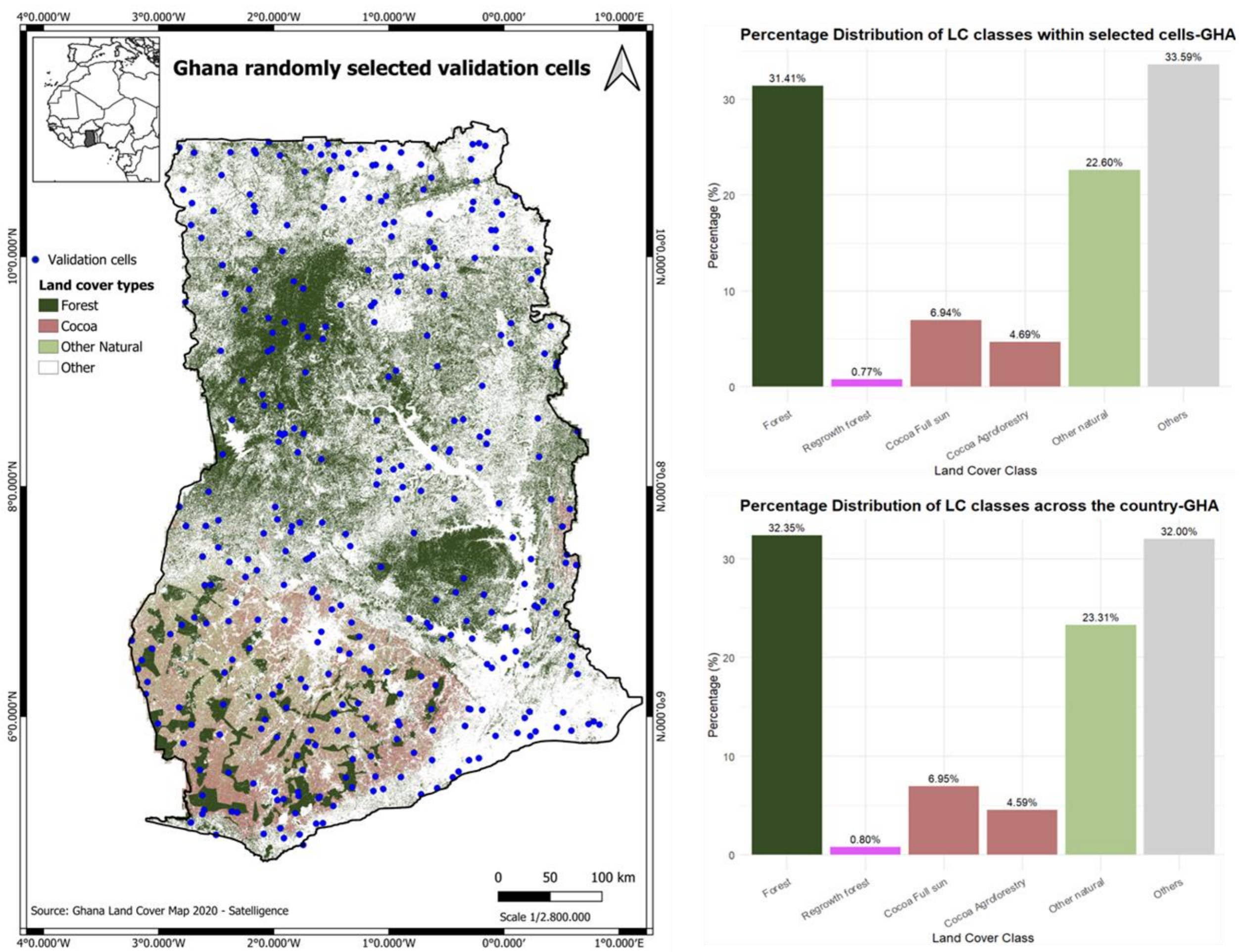

Figure 3 Validation sites in Ghana and their respective class distribution

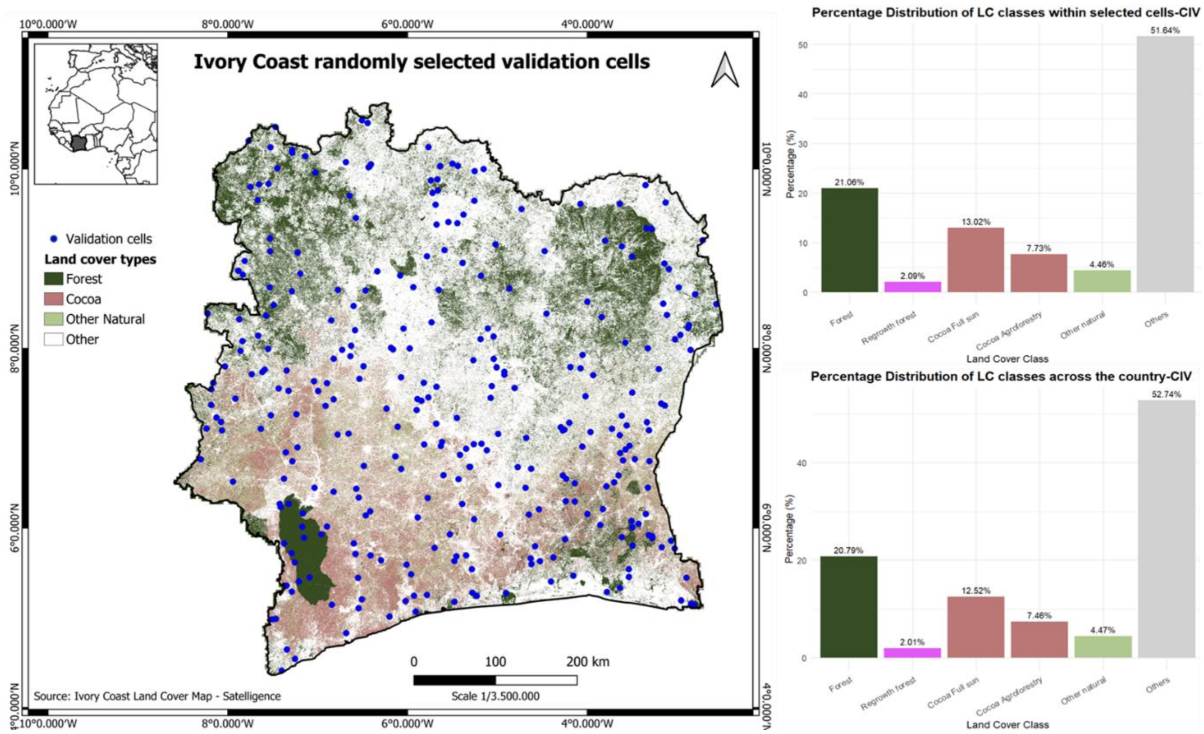

Figure 4 Validation sites in Cote D'Ivoire and their respective class distribution

As shown in Figure 3 and Figure 4, the stratified random sampling process yielded a total of 680 sites for validation by interpreters. The distribution of key land cover categories within the validation areas, derived from Satelligence data, closely mirrored the country-level distribution, also based on Satelligence data, with differences not exceeding 5%. This alignment ensures a representative and robust sampling framework for the analysis.

Cocoa

As shown in Figure 5, all datasets analyzed for their accuracy at identifying cocoa achieved a precision above 70% in Côte d'Ivoire. The FDaP and ETH-Cocoa datasets demonstrated the highest precision, with scores of 96% ± 4% and 91% ± 7%, respectively. In comparison, the Satelligence datasets and official national data showed lower precision, with scores of 75% ± 9% and 72% ± 14%, respectively. For recall, Satelligence outperformed the other datasets, achieving the highest score of 90% ± 11%. The ETH-Cocoa, FDaP, and official national data datasets followed with recall scores of 80% ± 12%, 44% ± 12%, and 41% ± 12%, respectively.

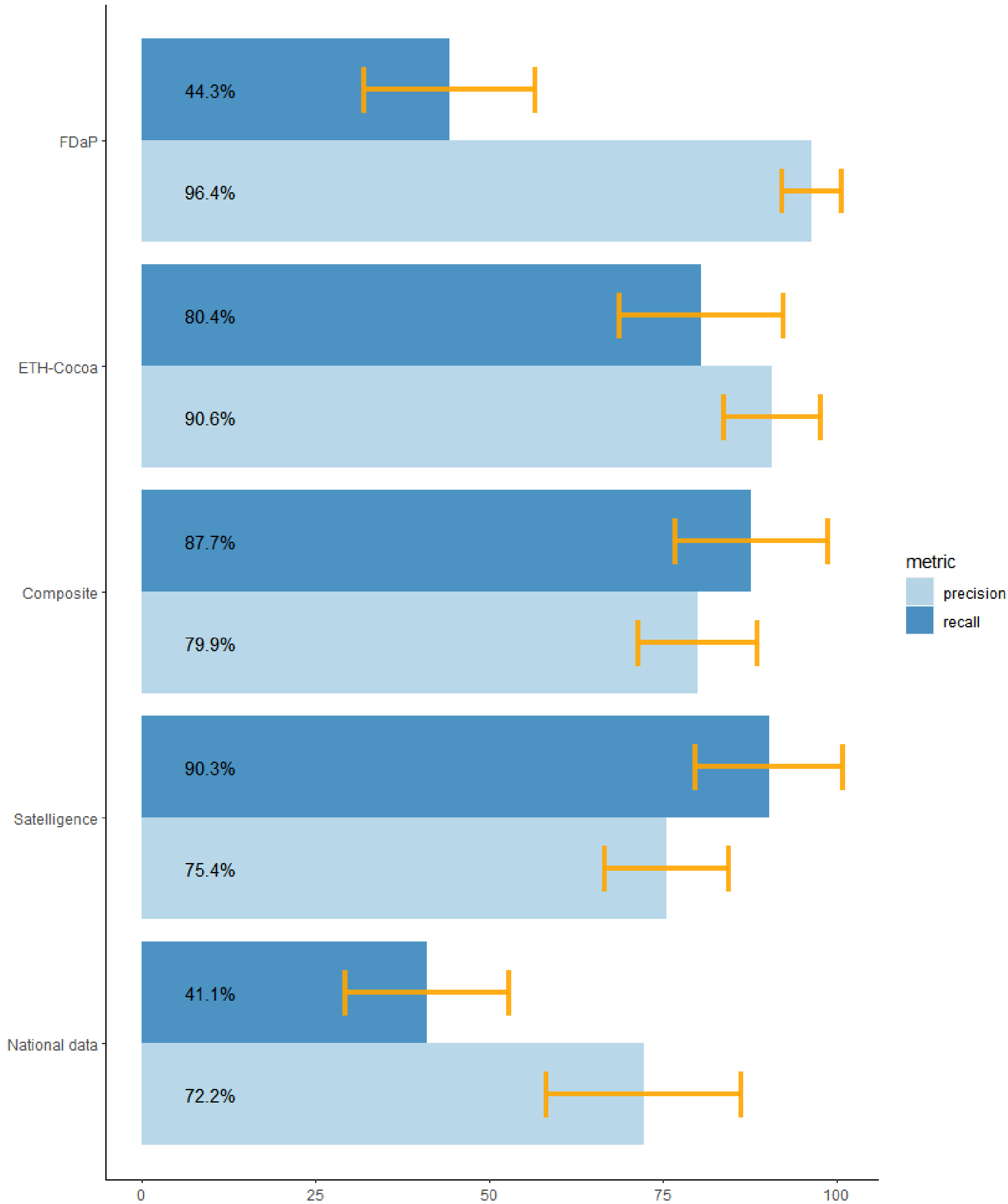

Figure 5 Precision and recall for cocoa in CIV

As shown in Figure 6, in Ghana, all datasets demonstrated good precision scores, with FDaP and official national data achieving the highest precision at 83.0% ± 12.2% and 82% ± 15%, respectively. The ETH-Cocoa and Satelligence datasets followed closely, with precision scores of 79% ± 8% and 78% ± 8%, respectively.

For recall, Satelligence obtained the highest score, reaching 87% ± 13%, closely followed by the ETH-Cocoa dataset with 86% ± 13%. The official national data achieved a moderate recall of 64% ± 12%, while FDaP recorded the lowest recall at 25% ± 10%.

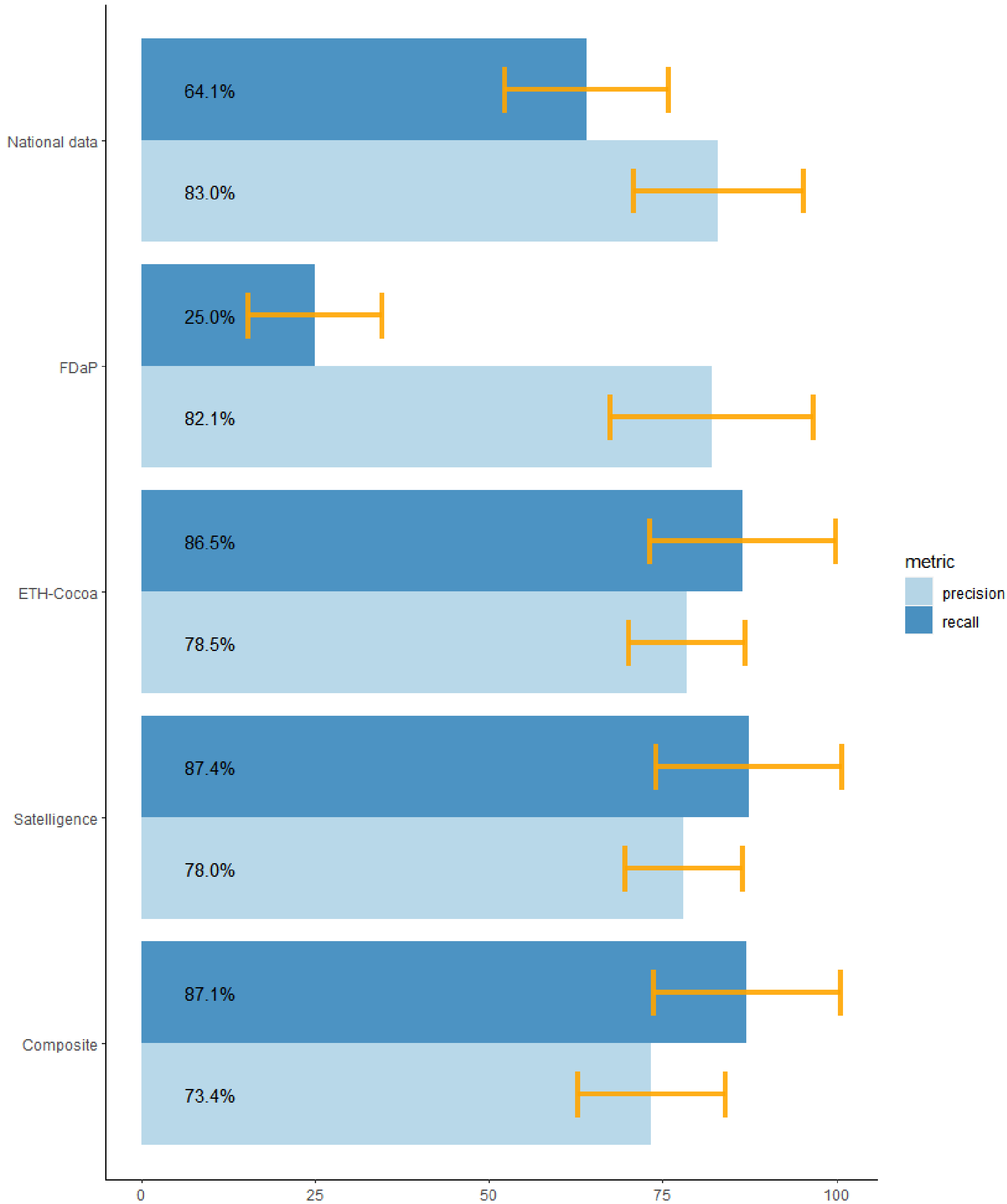

Figure 6 Precision and recall for cocoa in GH

Forests

As shown in Figure 7, for forest identification in Côte d'Ivoire, Satelligence and the locally developed datasets demonstrated significantly higher precision compared to global datasets. Satelligence achieved the highest precision at 87% ± 6%, followed by the official Côte d'Ivoire dataset with 73% ± 10%. Among global forest maps, the JRC-EU-V2 dataset had a lower precision of 64% ± 12%, while the GLAD-Forest dataset displayed a significant imbalance, with a notably low precision of 39% ± 7%. Tree cover maps performed even worse in terms of precision, with Hansen-30% scoring 35% ± 6%, ESA- WOC achieving 36% ± 8%, and DW recording 36% ± 7%. These results indicate that while global datasets and tree cover maps capture forested areas, they often misclassify other land types as forests, leading to lower precision.

In terms of recall of country level datasets, which measures the ability to detect all forested areas, the official Côte d'Ivoire dataset achieved a recall at 81% ± 12%, followed by Satelligence at 71% ± 13%. Global forest maps, despite their lower precision, generally exhibited higher recall. The JRC-EU-V2 dataset reached 80% ± 12%, while GLAD-Forest had the highest recall at 95% ± 7%, though at the cost of very low precision. Tree cover maps followed a similar pattern, with DW achieving 97% ± 3% recall and GFC-10% reaching 95% ± 7%, indicating their ability to detect tree-covered areas broadly. However, GFC-30%, which uses a stricter threshold, had a lower recall of 56% ± 13%, as it fails to capture open forests that fall under the FAO 10% canopy cover definition. This highlights a trade-off between precision and recall, where tree cover maps and global datasets tend to overestimate forest presence, while locally developed datasets provide more balanced results.

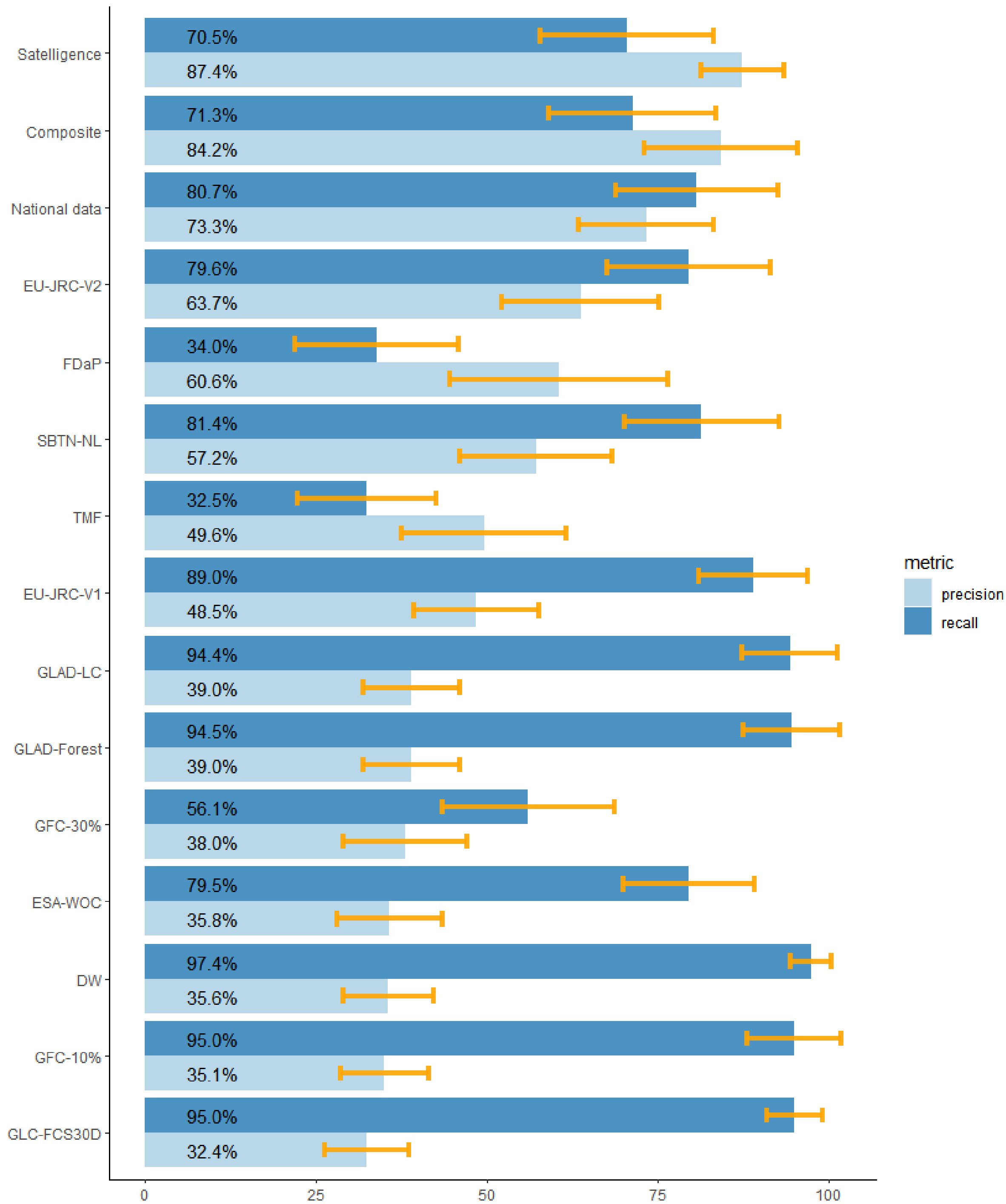

Figure 7 Precision and recall for forest in CIV

In Ghana, the highest precision scores were observed in locally tailored datasets and select global forest datasets. s-FDaP, SBTN-NL, and the official national dataset performed best, achieving precision values of 68% ± 18%, 66% ± 11%, and 63% ± 13%, respectively. Among global datasets, JRC-EU-V2 demonstrated a relatively good precision of 63% ± 11%, whereas GLAD-Forest and JRC-

EU-V1 had lower precision at 53% ± 9% and 52% ± 9%, respectively. Tree cover datasets, such as DW (43% ± 8%) and ESA-WOC (41% ± 10%), had the lowest precision.

In terms of recall, SBTN-NL excelled at 87% ± 8%, followed by the official Ghana dataset at 63% ± 13%. The JRC-EU-V2 dataset also performed well, achieving 88% ± 7% recall, while GLAD (94% ± 6%) and JRC-EU-V1 (95% ± 5%) exhibited extremely high recall, meaning they effectively captured most forested areas but at the cost of lower precision. In contrast, tree cover datasets such as DW and ESA-WOC demonstrated inconsistent performance, pairing low precision with variable recall.

Satelligence did not achieve strong results country-wide in Ghana, with a precision of 59% ± 11% and a recall of 83% ± 10%. Further inspection revealed that these low scores were primarily due to an overestimation of forests in savanna lands, where areas with sparse trees under 5 meters were misclassified as forest, contributing to over 60% of the errors.

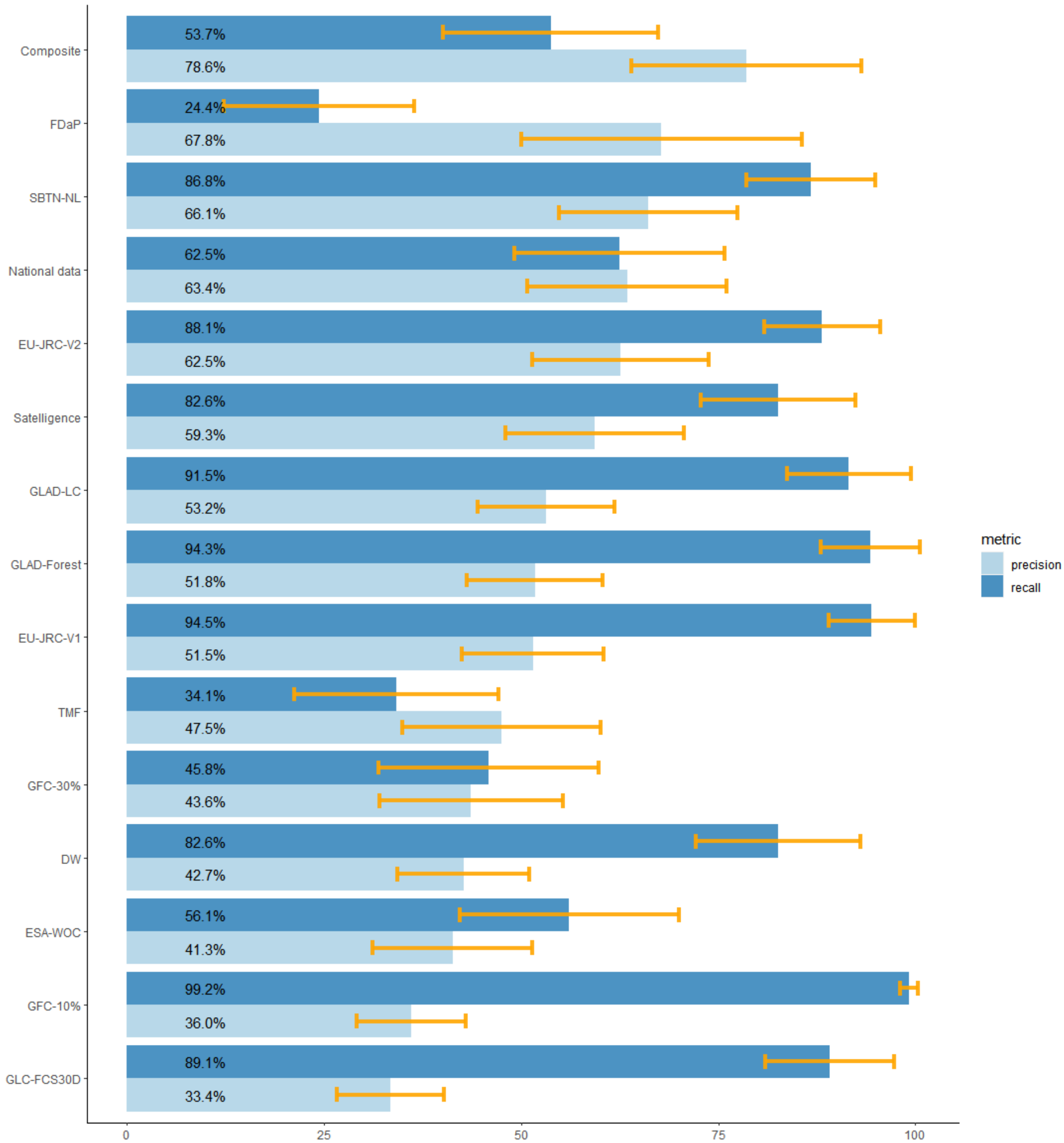

Figure 8 Precision and recall for forest in GH

Non-forest natural lands

For non-forest natural lands, all datasets achieved relatively low scores, suggesting that these land cover classes may not have been prioritized during dataset development and that further research is needed to improve their mapping accuracy. Among the tested datasets, the most balanced performance was observed in Ghana with Satelligence, which achieved a precision of 60% ± 14% and a recall of 63% ± 12%.

In comparison, most other datasets scored in the low 40% range for both precision and recall in both Côte d'Ivoire and Ghana, indicating significant challenges in accurately identifying and classifying non-forest natural lands across these regions. These results emphasize the need for enhanced methodologies and data tailored to these often-overlooked land cover types.

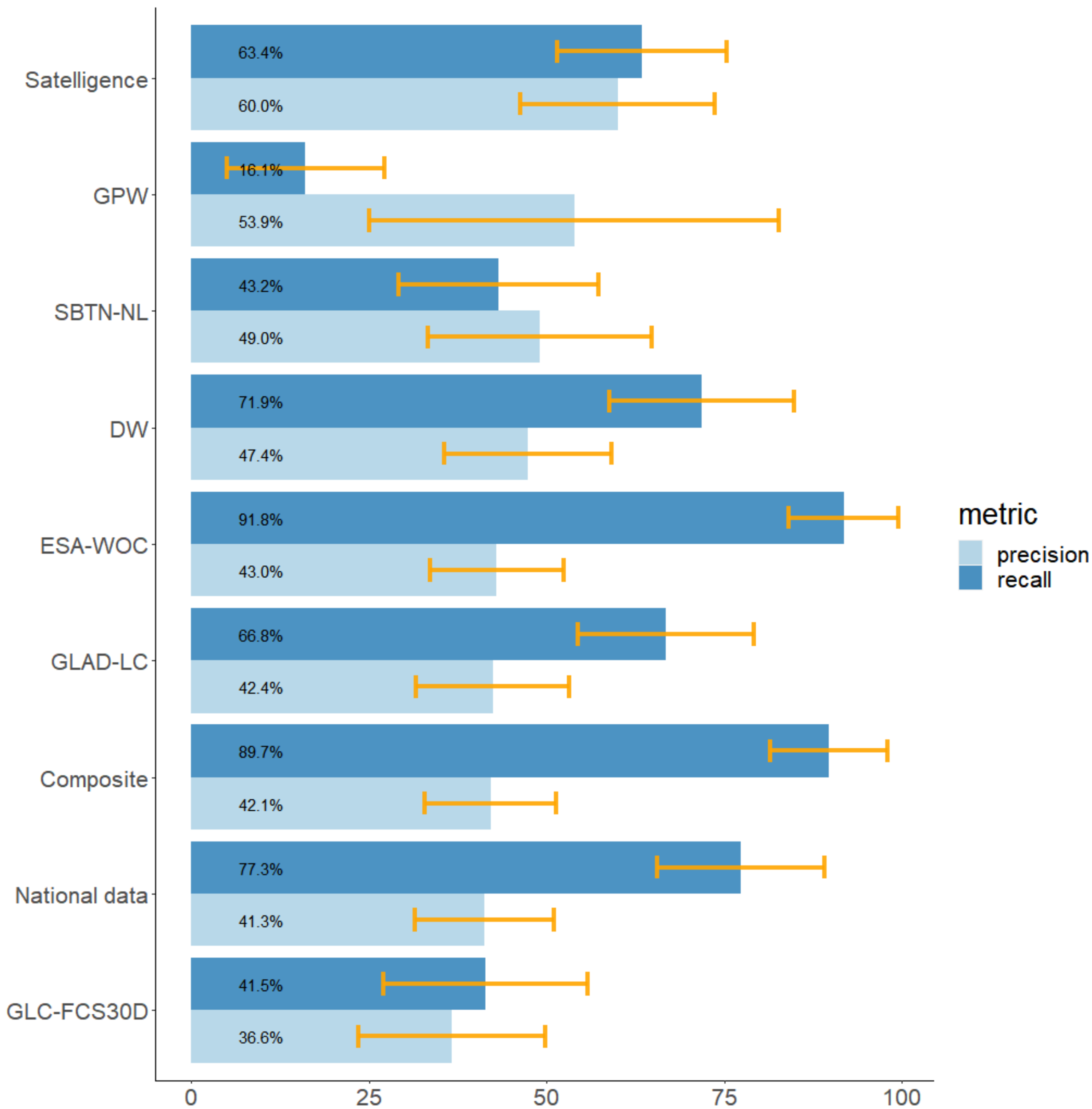

Figure 9 Precision and recall for non-forest natural lands in GH

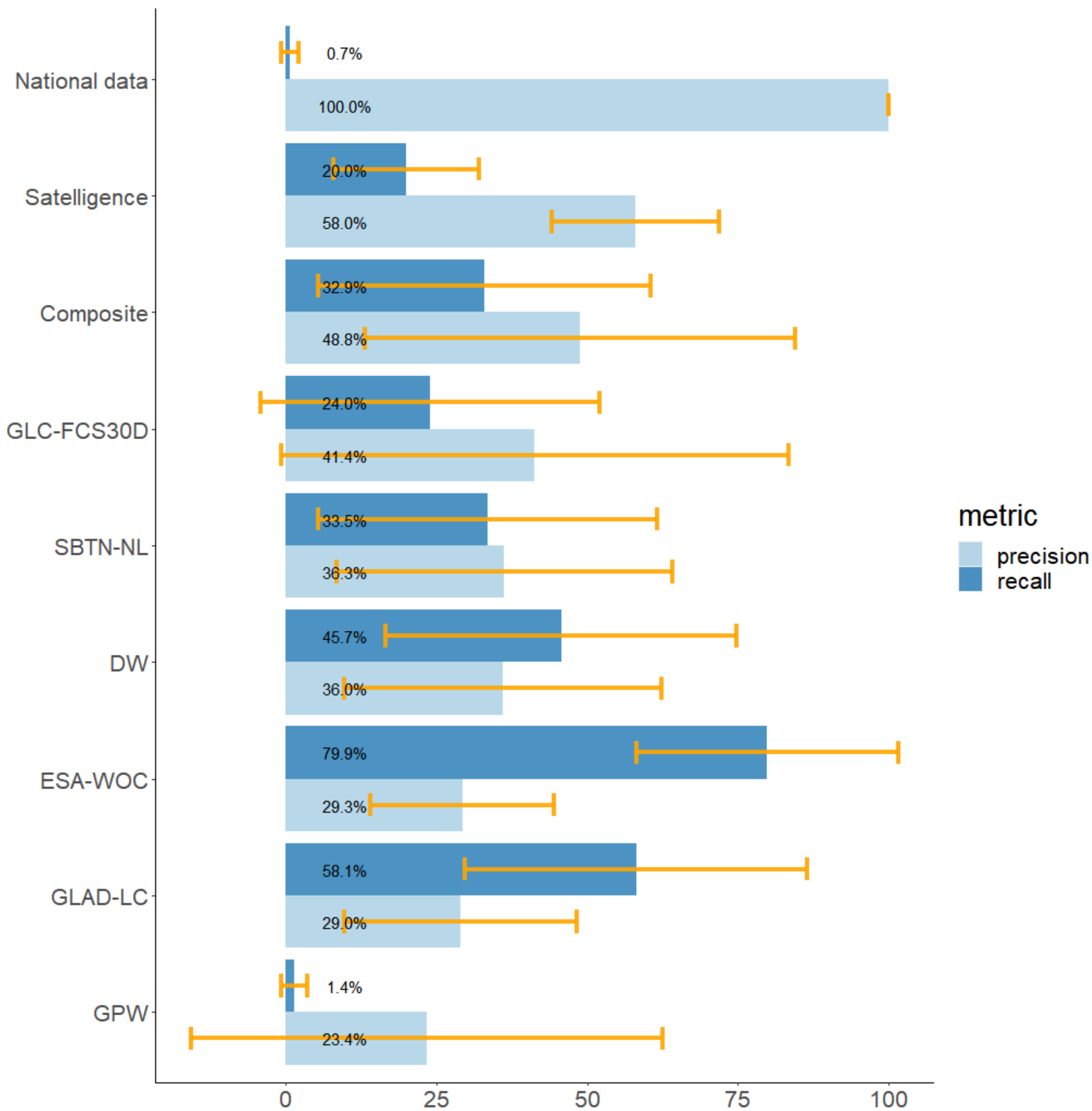

Figure 10 Precision and recall for non-forest natural lands in CIV

Composite Analysis

Figure 11 illustrates the decision tree used to construct the composite land cover dataset. This rule-based approach follows a sequential, hierarchical structure that prioritizes precision first and then progressively increases recall to enhance overall coverage. The tree reads from top to bottom, with each pixel being assigned a land cover category based on the most reliable datasets available. Once a pixel is assigned a value, it remains fixed throughout the process, ensuring consistency in classification.

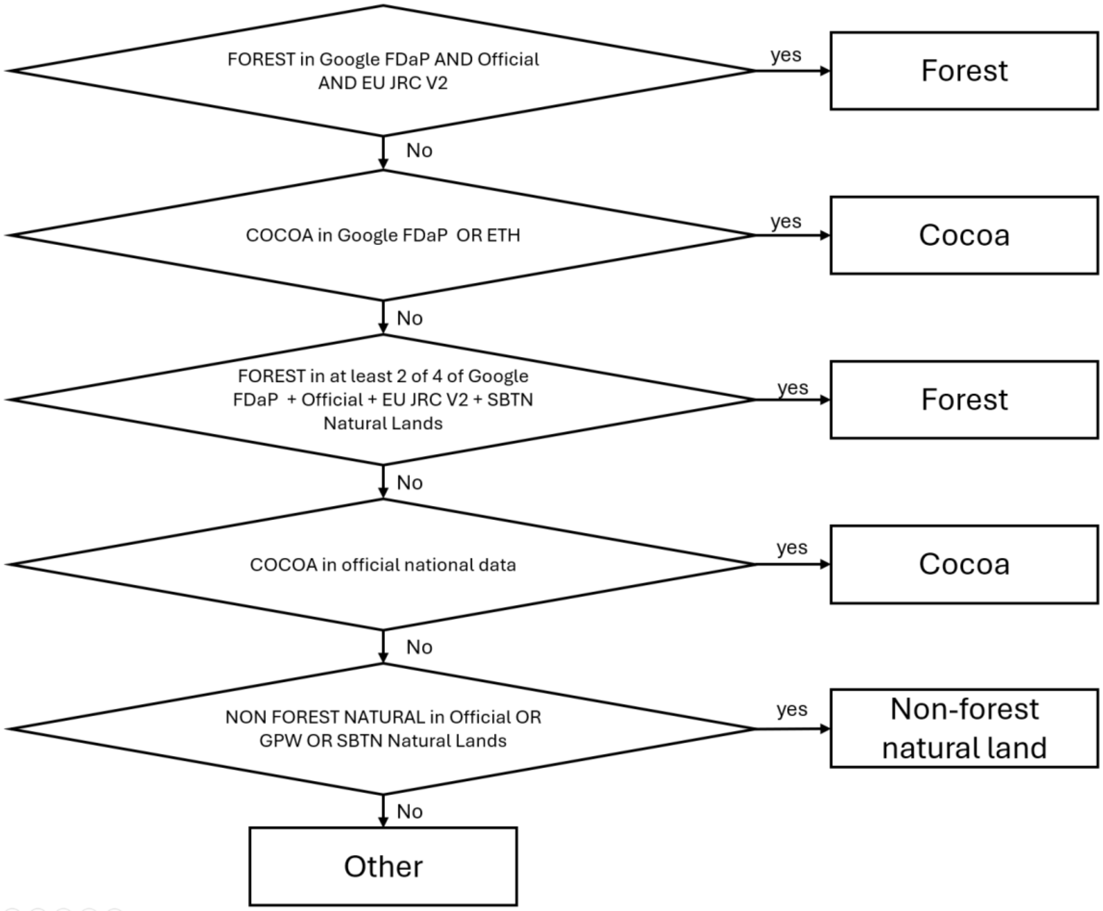

Figure 11 Decision tree to allocate a class to a given pixel.

Composite Map Construction Logic

1. Forest Layer (High Confidence Zones) – The process begins by assigning forest classification to pixels where there is agreement among the three best-performing datasets in terms of precision. This ensures that the most reliable forest pixels are established first.

2. Cocoa Layer (High Precision Priority) – Cocoa pixels are then assigned using the FDaP and ETH-Cocoa datasets, both of which have demonstrated very high precision in cocoa classification.

3. Expanding Forest Recall – To enhance forest coverage while maintaining precision, additional forest pixels are added based on a less restrictive rule, improving recall without significantly increasing false positives.

4. Expanding Cocoa Recall – Additional cocoa pixels from official datasets are incorporated to further improve the completeness of the cocoa classification.

5. Natural Non-Forest Layer (Least Reliable Class) – Finally, non-forest natural ecosystems are assigned, recognizing that this category has the lowest precision and recall across all maps.

By structuring the process in this way, we ensure that high-confidence forest pixels are prioritized, followed by high-precision cocoa pixels, and finally, an increase in recall for both forest and cocoa before assigning the least reliable land cover category.

The composite approach yielded high accuracy, delivering consistent and balanced results across land cover categories in both Côte d'Ivoire and Ghana. In Côte d'Ivoire, the composite approach achieved an overall accuracy of 75% ± 7%, with particularly strong performance for Cocoa (precision: 80% ± 9%, recall: 88% ± 11%) and Forest (precision: 84% ± 11%, recall: 71% ± 12%). The Other category also performed well with a precision of 72% ± 11% and a recall of 81% ± 10%, though non-forest natural lands remained challenging, scoring a precision of 49% ± 36% and a recall of 33% ± 28%.

In Ghana, the composite approach achieved an overall accuracy of 60% ± 7%, demonstrating balanced results across categories. The cocoa category performed strongly with a precision of 73% ± 11% and a recall of 87% ± 14%, while Forest attained a precision of 79% ± 15% and a recall of 54% ± 14%. The Other category achieved a precision of 79% ± 12%, but its recall was lower at 39% ± 11%. Conversely, non-forest natural lands had a high recall of 90% ± 8% but lower precision at 42% ± 9%.

The composite approach outperformed any single publicly available dataset in forest precision and achieved the highest recall for cocoa. It performed less well for non-forest natural areas (7th in Côte d'Ivoire, 3rd in Ghana in term of precision), as the decision tree was not optimized for that category.

Full Criteria Assessment

Table 2 and 3 present the precision and recall for each land cover dataset, along with scores for completeness, data management, inclusiveness, and the resulting final score. In the background, precision and recall are combined using the F-score, which favors datasets with balanced performance. This value is then weighted twice as much as the other criteria in calculating the final score.

Table 2 Overall assessment scores for land cover datasets in CIV

	Precision	Recall	Completeness	Data Management	Inclusiveness	Final Score
				Forest		
Satelligence	87%	70%	70%	65%	66%	73%
Official-CIV	73%	81%	46%	65%	22%	63%
EU-JRC-V2	64%	80%	53%	100%	0%	62%
EU-JRC-V1	48%	89%	53%	100%	0%	57%
DW	36%	97%	57%	100%	0%	52%
GFC-10%	35%	95%	53%	100%	0%	51%
FDaP	61%	34%	51%	98%	33%	51%
ESA-WOC	36%	80%	50%	100%	0%	50%
GFC-30%	38%	56%	53%	100%	0%	48%

GLC	32%	95%	38%	100%	0%	47%
TMF	50%	33%	49%	100%	0%	44%
GLAD-Forest	39%	52%	49%	22%	0%	36%
			Cocoa			
Satelligence	75%	90%	70%	65%	66%	75%
ETH-Cocoa	91%	80%	42%	100%	0%	69%
FDaP	96%	44%	51%	98%	33%	60%
Official-CIV	72%	41%	46%	65%	22%	49%

Table 3 Overall assessment scores for land cover datasets in GH

	Precision	Recall	Completeness	Data Management	Inclusiveness	Final Score
			Forest			
Satelligence	59%	83%	70%	65%	66%	68%
EU-JRC-V2	63%	88%	53%	100%	0%	64%
EU-JRC-V1	51%	94%	53%	100%	0%	60%
DW	43%	83%	57%	100%	0%	55%
GFC-10%	36%	99%	53%	100%	0%	52%
Official-GH	63%	63%	46%	25%	22%	49%
ESA-WOC	41%	56%	50%	100%	0%	48%
GLAD-Forest	52%	94%	49%	22%	0%	48%
GFC-30%	44%	46%	53%	100%	0%	48%
GLC	33%	89%	38%	100%	0%	47%
FDaP	68%	24%	51%	98%	0%	42%
TMF	47%	34%	49%	0%	0%	30%
			Cocoa			
Satelligence	73%	87%	70%	65%	66%	74%
ETH-Cocoa	78%	87%	42%	100%	0%	67%
Official-GH	83%	64%	46%	25%	22%	55%
FDaP - Cocoa	82%	25%	51%	98%	0%	43%

Discussion

All datasets evaluated in this study were tested using a national-level geographic scope and based on a strict application of the FAO forest definition. However, it is important to note that several of these datasets were not originally designed for this specific scope or definition. For example, Satelligence data were developed and optimized primarily for the cocoa-growing regions of Ghana, rather than for national-scale application. Similarly, the FDaP forest layer was designed to identify naturally regenerating forests with minimal signs of human disturbance, which excludes certain forest types that still fall within the FAO definition. In addition, datasets based on tree cover, such as GFC and DW, were not designed to differentiate between natural forests and tree crops, leading to potential misclassification in areas with agricultural tree cover.

While some datasets were originally developed for different geographic scopes or definitions, the accuracy results presented in this study are robust and valid for assessing land cover, in particular forests and cocoas areas, as defined within the scope and methodology of this study. These results may not reflect how the same datasets perform in other contexts or regions.

Strengths and Limitations of Global Datasets

Global land cover datasets, while widely available and well-documented, exhibit significant limitations in the context of deforestation-free compliance, particularly in complex landscapes like Ghana and Côte d'Ivoire. These datasets are unable to effectively differentiate between forests and cocoa systems, which compromises their applicability for monitoring agroforestry, cocoa-driven deforestation, and cocoa driven LUC for GHG accounting.

Despite these limitations, global datasets offer strengths in other areas. Their comprehensive documentation and accessibility make them transparent and easy to integrate into analyses. Furthermore, the use of consistent methodologies across regions ensures that the data is comparable, enabling streamlined integration into global assessments. While these features make global datasets useful for broad-scale applications, they fall short of meeting the requirements of localized deforestation-free compliance.

Strengths and Limitations of Local Datasets

Locally tailored datasets demonstrate significantly higher precision compared to global datasets, making them more suitable for deforestation-free compliance. Official national datasets, while slightly less accurate than commercial datasets, provide robust and comprehensive data, offering a valuable public resource. However, these datasets often lack proper documentation, metadata, and accessible data management practices, which can limit their usability and transparency.

Emerging Alternatives

Emerging datasets, such as those developed by the Forest Data Partnership, represent promising alternatives. These datasets have shown high precision but low recall, often underestimating areas such as planted cocoa systems. While they are still evolving, future iterations have the potential to become more reliable as they integrate more inclusive and representative data.

Commercial Solutions

As shown in Table 2 the service provided by Satelligence currently offers the most robust and comprehensive land cover maps. These solutions are good in documentation and metadata, but their resources are hosted exclusively on proprietary platforms, which may limit long-term accessibility. Additionally, while Satelligence integrates local data to improve relevance, a formal inclusiveness strategy for engaging diverse stakeholders is currently lacking.

As mentioned initially, Satelligence was included in this study to support ongoing WCF led work and to provide participating WCF members with the quality assurance needed to confidently use their analyses and service. Other commercial providers have not been assessed in this study, but they may also have highly accurate maps; it is recommended that 3rd party accuracy assessments are conducted by all commercial providers and shared publicly so that interested parties can make informed decisions on which data and service to use.

To enhance their usability and accessibility, it is recommended that commercial solutions publish metadata and documentation in public repositories. Additionally, private service providers should be encouraged to establish commercial agreements with platforms to improve data availability (e.g., Satelligence on GFW). This would facilitate broader access to high-quality datasets while ensuring their integration into widely used monitoring frameworks. Also recommended that commercial solutions have third party assessments of accuracy that are made publicly available. Moreover, developing a structured inclusiveness strategy would further strengthen the credibility and applicability of commercial datasets in diverse contexts. Despite these challenges, commercial solutions remain a valuable resource for land cover monitoring.

Composite Analysis

The composite approach demonstrated high overall accuracy and balanced performance across land cover categories in Côte d'Ivoire and Ghana, with strong results for Cocoa (Côte d'Ivoire: precision 80% ± 9%, recall 88% ± 11%; Ghana: precision 73% ± 11%, recall 87% ± 14%) and Forest (Côte d'Ivoire: precision 84% ± 11%, recall 71% ± 12%; Ghana: precision 79% ± 15%, recall 54% ± 14%). However, non-forest natural lands showed lower precision and recall, highlighting classification challenges in those categories.

While the precision and recall for individual categories are not the highest obtained across all tested methods, this composite approach provides a balanced and consistent performance across all land cover types. These results highlight the value of integrating multiple datasets in a composite framework, offering a practical and reliable solution for achieving robust land cover mapping using heterogeneous data sources.

However, it is important to note that the datasets used in this composite approach were selected based on the results of the validation process. As a result, the independence of the test sampling is compromised, which may introduce bias and artificially inflate the accuracy estimates. This limitation should be considered when interpreting the results of this analysis.

Limitations

Snapshot in Time

This study validates land cover datasets for the year 2020, offering a single-year assessment based on a specific cut-off date, which does not fully capture temporal land cover changes. However, change detection methods achieve 80–95% precision in identifying deviations from the baseline. Most errors, particularly false positives, originate from inaccuracies in the baseline data, such as misclassifying tree crops as forests, rather than flaws in the detection algorithms. However, some coarser-resolution products, such as GFC with 30m resolution, have higher omission rates for small-scale deforestation events (sub-pixel size), which are often associated with smallholder-driven deforestation. Given that this study was conducted for a single year, this effect could not be assessed. Consequently, the reliability of detected changes is closely tied to the quality of the baseline data. While 2020 is not expected to be an outlier in terms of data quality, datasets covering additional years using a consistent methodology are expected to yield comparable results.

Composite Approach Considerations

The validation process used the same dataset for both selection and accuracy assessment, which may introduce bias and potentially inflate performance estimates. While the composite approach improves consistency across different land cover types, further research is needed to refine these methods and validate results independently. Initial findings indicate promise, but additional evaluations are required to enhance dataset integration and ensure robustness across varying landscapes and conditions.

Carbon and Biomass

The Role of Remote Sensing in Carbon Removal Assessment

Under the draft GHG Protocol, reporting on carbon removals requires the use of primary, site-specific data derived directly from a company's operations or value chain. Remote sensing can support this process, but only when calibrated with ground-based measurements such as inventory plots or repeated carbon stock sampling. High-resolution technologies like LiDAR or UAV-based photogrammetry can provide detailed insights into above-ground biomass, but their outputs must be validated with field data to meet reporting standards. These methods may be considered equivalent to direct measurements when properly calibrated, though they are not yet fully applicable to all carbon pools, such as soil carbon. To comply with the GHG Protocol, removal quantification must use Tier 2 or Tier 3 methods, incorporating empirical data and statistically sound uncertainty estimates (WRI & WBCSD, 2022).

Given these requirements, no freely available remote sensing datasets currently meet the precision and calibration standards needed for carbon removal assessments. Commercial solutions, which integrate drone-based surveys, airborne LiDAR, and machine learning models trained on high-resolution imagery and field data, offer more details analysis. These methods were not assessed within this study.

Public datasets, including tree canopy height layers discussed in the Tree Planting section, may support landscape-scale monitoring, but lack the granularity and local calibration needed for accurate removal estimation at the farm level. For these reasons, this section focuses on emissions from deforestation, rather than in-farm carbon removals. Further guidance on GHG accounting in cocoa supply chains is available in the GHG Accounting Manual for Cocoa, developed by Quantis and WCF (Rizzo et al., 2025).

Selected datasets

Global Forest Carbon Fluxes (2001–2023)

The GFCF dataset (Harris et al., 2021) provides a comprehensive evaluation of net forest carbon fluxes by calculating the balance between carbon emissions and removals across global forests. Using IPCC guidelines, this dataset includes detailed layers on cumulative carbon emissions from forest disturbances, carbon removals through forest regrowth, and net carbon exchange.

Satelligence

The Satelligence platform integrates 20 years of data with its land cover datasets to support Scope 3 carbon reporting using linear discounting. Its aboveground biomass modeling incorporates data from multiple satellite sources, including Sentinel-1, Sentinel-2, Landsat, ALOS PALSAR, GEDI, and ICESAT. This approach is occasionally validated with airborne LiDAR and field measurements, ensuring robust and precise carbon accounting insights.

FCPF Carbon Fund – Taï National Park Emission Reductions Project

This project supports emission reduction assessments through a robust, sampling-based approach for land cover change detection. It combines hybrid machine learning with human visual interpretation, leveraging diverse satellite imagery such as ESA CCI and Landsat, alongside multiple change detection algorithms including BFAST, CUSUM, CCDC, and LandTrendR. Rigorous cross-validation techniques and strict quality control processes ensure high accuracy and minimized bias in results. The systematic 1km x 1km sampling grid enables consistent, long-term monitoring of land cover changes with a relative sampling error of less than 15%, providing a reliable framework for emission reduction and biomass assessments.

Method Overview

A quantitative assessment of the datasets for carbon and biomass monitoring was not conducted in this study; an evaluation would require extensive and resource-intensive fieldwork beyond the scope of this analysis. Instead, to evaluate the **usability** of each dataset within the context of cocoa supply chain Scope-3 reporting, we conducted a structured assessment of their strengths and weaknesses. This evaluation was based on a systematic review of scientific literature, focusing on three key aspects. First, we examined validation quality, assessing whether the dataset was validated using appropriate methods and if the results demonstrated robust accuracy. Second, we analyzed methodological limitations, identifying any constraints in the dataset's creation that could impact its suitability for carbon monitoring. Lastly, we reviewed reported limitations, considering documented constraints that might affect the dataset's intended application. This structured approach ensured a comprehensive assessment of each dataset's reliability and relevance for Scope-3 reporting in the cocoa sector.

In addition to evaluating datasets usability for carbon and biomass monitoring, we conducted a systematic review of metadata and documentation to assess completeness and data management. **Completeness** examines whether the datasets and platforms provide adequate geographic coverage, temporal resolution, and data granularity. It also considers if historical data is sufficient to monitor GHG emissions. **Data management** assesses the quality of metadata and data management practices, ensuring adherence to the FAIR principles (Findable, Accessible, Interoperable, and Reusable). This evaluation considers the clarity, completeness, and accessibility of dataset documentation.

Overall Assessment

The overall assessment of carbon datasets is presented in Table 4. Satelligence achieved the highest final score of 64%, with balanced performance in usability (66%), completeness (66%), and data management (57%), reflecting its robustness as a tailored carbon monitoring solution. The Global Forest Carbon Fluxes (GFCF) dataset followed with a final score of 59%, excelling in data management (96%) and completeness (73%), but scoring lower in usability (33%), which highlights its reliance on suboptimal baselines for complex landscapes. Satelligence and GFCF achieved similar overall scores, but for different reasons. Satelligence demonstrated strength in providing a more reliable option for carbon analysis, while GFCF remains a solid publicly available alternative, particularly for applications that require strong data management practices. However, GFCF relies on global datasets, which lack the precision needed for detailed local analysis, making it less suitable for region-specific assessments.

Table 4. Overall assessment scores for carbon datasets

	Usability	***Completeness***	***Data Management***	***Final Score***
Satelligence	Good	66%	57%	64%
GFCF	Limited	73%	96%	59%

Since the FCPF Carbon Fund – Taï National Park Emission Reductions Project is a method rather than a standalone dataset and has only been applied at a sub-country level in Côte d'Ivoire, it could not be fully assessed using the same criteria as the other datasets. As a result, it is not included in Table 4.

Global Forest Carbon Fluxes (2001–2023)

Self-reported validation

The validation of the GFCF results was conducted through a comparative analysis with the Global Carbon Project (GCP), a globally recognized initiative that compiles and harmonizes carbon cycle data from diverse scientific studies. The Global Carbon Project aggregates findings from various sources to establish a comprehensive and standardized understanding of the Earth's carbon fluxes.

To assess the accuracy and consistency of the GFCF findings, the authors compared their globally aggregated results against the GCP's global carbon budget (2001–2018). This comparison provided insights into carbon sources and sinks, with a particular focus on land-use change emissions, and forest carbon fluxes. While the two datasets are not directly comparable due to differences in scope, data sources, methodologies, and reporting structures, the analysis highlighted general alignment in key trends and totals. Specifically, total carbon sources were estimated at 42.0 $GtCO_2\ yr^{-1}$ in this study, compared to 37.3 $GtCO_2\ yr^{-1}$ in GCP (+12.6%), while total carbon sinks were 41.4 $GtCO_2\ yr^{-1}$ in this study, versus 36.1 $GtCO_2\ yr^{-1}$ in GCP (+14.7%).

Additionally, the authors conducted sensitivity analyses to evaluate how different data sources and model assumptions influenced overall GHG flux estimates at various spatial scales. This approach helped quantify uncertainties and assess the robustness of the study's carbon flux estimates. The full results of the sensitivity analysis are provided as supplementary material of the publication[1].

Warning: All validation and sensitivity analyses were conducted at global and regional scales, meaning that these results may not necessarily hold at the local scale. As shown in our analysis, accuracy and consistency vary significantly when applied to specific landscapes, such as cocoa-growing regions in Côte d'Ivoire and Ghana. Therefore, caution should be exercised when extrapolating these findings to local contexts.

Discussion

Usability

The GFCF is the most complete publicly available resource for carbon flux analysis. Its broad scope and accessibility make it a preferred choice for studies that rely solely on public data sources, providing valuable insights into carbon emissions, removals, and net fluxes across global forests.

The dataset relies heavily on GFC datasets (Hansen et al., 2013)as its forest cover and loss baseline. While this baseline provides global consistency, it has proven suboptimal in complex and fragmented landscapes, such as cocoa-growing regions in Côte d'Ivoire and Ghana. This limitation necessitates caution when using the carbon flux dataset in these specific contexts, as it may not accurately capture the complex dynamics of change within these landscapes.

A notable limitation of the dataset is its lack of year-specific data for forest loss or gain, requiring users to integrate external sources such as (Hansen et al., 2013) for temporal specificity in change detection. However, given the identified limitations of this dataset in Côte d'Ivoire and Ghana, this integration must be approached carefully to avoid inaccuracies in year-specific assessments.

Other forest datasets may also be used, but to be suitable, they must offer a long and consistent time-series of deforestation detection, ideally spanning at least 20 years. The key challenge in developing such datasets is establishing a reliable baseline of forest cover from two decades ago, a period when high-resolution imagery was scarce and modern publicly available satellite datasets such as Landsat 8 and 9, Sentinel-1, and Sentinel-2 did not yet exist. As seen in this study, defining a forest baseline for 2020 in critical cocoa landscapes already presents significant challenges, underscoring how difficult it is to construct a robust and consistent long-term deforestation dataset.

Currently, only a few publicly available datasets provide such historical deforestation time series, including Hansen et al. (2013), TMF, and GLAD. As with land cover analysis, a convergence of

[1] Available here : https://static-content.springer.com/esm/art%3A10.1038%2Fs41558-020-00976-6/MediaObjects/41558_2020_976_MOESM1_ESM.pdf

evidence approach may be the most appropriate, combining results from Hansen et al. and TMF while thoroughly assessing uncertainties in the outputs. However, from our current knowledge, such a method has not been extensively tested and would require further research to validate its effectiveness.

While some official historical land cover datasets exist, they often lack the consistency and robustness of newer datasets, limiting their utility for long-term carbon flux analysis. To address this gap, countries are encouraged to develop and provide high-quality historical data using consistent methodologies. Such efforts would significantly enhance the reliability of long-term carbon assessments and improve the dataset's applicability for historical trend analysis.

Overall, the GFCF dataset serves as a valuable dataset for carbon flux analysis on a global scale but requires careful consideration of its limitations and integration with complementary datasets for site-specific assessments.

Completeness

GFCF dataset offers strong capabilities in geographic and temporal coverage, making it a valuable resource for carbon emissions assessments over time. It provides excellent global coverage, ensuring its relevance for monitoring the cocoa value chain across different regions. Additionally, its temporal scope is excellent, offering sufficient historical data (over 20 years) to support GHG assessments. The dataset is also well-documented, with its methodology published in a peer-reviewed journal, ensuring transparency and credibility.

However, the dataset has limitations in its spatial resolution, definitions, and maintenance reliability. Its limited spatial resolution restricts its ability to precisely analyze cocoa landscapes at the plot level, which may affect its usability for localized land use assessments. Similarly, its land cover classifications do not fully align with FAO forest definitions. Its validation and uncertainty documentation are rated as good. Additionally, its communication about its maintenance reliability is limited, raising concerns about whether it will be regularly updated to support long-term monitoring efforts.

Data management

The GFCF dataset demonstrates strong data accessibility and management features, making it a well-structured resource for carbon flux assessments. It is hosted on a platform that allows access through an API, enabling seamless integration into analytical workflows. Additionally, it supports data downloads for offline analysis, providing flexibility for users working in different environments. The dataset is made available under a Creative Commons Attribution 4.0 International License, ensuring open access and ease of use. Furthermore, its metadata is published on Google Earth Engine, a widely recognized data platform that enhances discoverability and accessibility.

A key strength of this dataset is its detailed and well-documented metadata, which provides clear and comprehensive information on its methodology, structure, and limitations. This level of transparency significantly improves usability and reliability, allowing researchers and practitioners to make informed decisions when applying the dataset for carbon flux monitoring, deforestation analysis, and other environmental assessments.

Satelligence

Self-reported Validation

Satelligence validated its deforestation detection system through a series of case studies across multiple regions, comparing its accuracy and timeliness with other available systems. This validation process was conducted for both forests and agroforestry systems, ensuring applicability across different land cover types. The validation relied on visual interpretation of Sentinel-2 and Planet imagery, supplemented with ground truth data from field partners. This approach ensured that the detection system specifically identified deforestation rather than other forms of tree cover loss, such as plantation clearing.

Additionally, Satelligence's above-ground biomass modeling has been calibrated and validated using airborne LiDAR and field measurements, confirming the robustness and accuracy of its carbon stock estimations for both forests and agroforestry systems. A peer-reviewed publication, jointly submitted with the European Space Agency (ESA), presents the methodology and assessment results and is currently under review. However, until the publication is released, quantitative results cannot be publicly shared. While these results have been reviewed as part of this study, they could not be formally included in this report.

Discussion

Usability

Satelligence provides a comprehensive and reliable Scope-3 reporting service, leveraging its accurate cover maps. This enables precise insights for carbon accounting making it a valuable tool for organizations seeking to meet sustainability targets and ensure transparency in supply chain reporting.

Satelligence's aboveground biomass modeling integrates diverse satellite data sources, including Sentinel-1, Sentinel-2, Landsat, ALOS PALSAR, GEDI, and ICESAT. This multi-sensor approach is occasionally validated with airborne LiDAR and field measurements, significantly enhancing the accuracy and reliability of biomass estimates. Such integration ensures robust data outputs for carbon monitoring and compliance purposes.

The platform's methodological documentation is only available upon request, while training materials and metadata are publicly accessible but hosted exclusively on Satelligence's platform. This limited accessibility poses challenges for long-term transparency and broader stakeholder engagement.

Overall, Satelligence offers a robust and reliable solution for carbon and biomass monitoring, with the integration of locally measured data leading to higher accuracy. Addressing its documentation accessibility would further enhance its utility and transparency for a wide range of users.

Completeness

The Satelligence carbon assessment dataset demonstrates strong completeness, particularly in spatial resolution, geographic coverage, temporal scope, and definitional accuracy. With excellent spatial resolution, it provides highly granular data suitable for analyzing cocoa landscapes at the plot level. Its geographic scope is also excellent, ensuring comprehensive coverage of regions relevant to monitoring the cocoa value chain globally. Additionally, the dataset has a robust temporal scope including over 20 years of historical data, making it well-suited for GHG assessments. The dataset adheres to credible classification standards, aligning with FAO forest definitions. Furthermore, maintenance reliability is excellent, indicating that the dataset is actively updated to support consistent monitoring efforts.

However, uncertainty documentation and validation remain limited, as comprehensive independent validation results are not yet publicly available. That said, a peer-reviewed publication, jointly submitted with the European Space Agency (ESA), presents the methodology and assessment results and is currently under review. Until the publication is released, quantitative validation results cannot be shared publicly. While these results have been reviewed as part of this study, they could not be formally included in this report.

Finally, historical impact is rated as good, with documented use in various applications related to deforestation monitoring and supply chain analysis.

Data management

The dataset is accessible via API, ensuring reliable data access and integration into analytical workflows. Additionally, it is provided with a clear data usage license, outlining the terms of use. However, several aspects of metadata quality and accessibility are limited or very limited, which affects its findability, interoperability, and long-term reusability.

While metadata are available, they lack a globally unique identifier, are not registered in a common searchable database, and do not clearly link to dataset versions. Additionally, offline availability is very restricted, and there is no clear strategy for long-term data retention, which raises concerns about its future accessibility.

FCPF Carbon Fund – Taï National Park Emission Reductions Project

Self-reported validation

While no validation per say was performed in this study, uncertainty was systematically documented and managed through a structured Quality Assurance/Quality Control (QA/QC) approach, addressing both systematic biases and random errors. The assessment focused on three key areas: activity data, emission factors, and model integration.

For activity data, measurement uncertainty primarily stemmed from land cover classification errors due to variations in color, texture, and seasonal influences in satellite imagery. To mitigate this, interpreter training, standardized protocols, and confusion matrix analyses were employed. Sampling uncertainty was addressed through a stratified probabilistic sampling approach, ensuring a balanced representation across different land cover types. Despite this, some rare land cover changes exhibited high variance, which remained a source of residual uncertainty.

In terms of emission factors, uncertainty was introduced through tree measurement errors, biomass estimation models, and default assumptions for wood density. While tree height and diameter measurements were subject to field verification and standardized protocols, biomass estimation relied on pantropical allometric equations due to the lack of species-specific models for Côte d'Ivoire. The use of default values for wood density in cases where species-specific data was unavailable introduced potential bias, despite cross-validation efforts.

Model integration uncertainty was managed by harmonizing activity data with emission factors to ensure consistency in carbon flux estimates. Error control mechanisms were embedded in emission and removal calculations to prevent double-counting and ensure methodological coherence. However, uncertainties persisted in the classification of agroforestry systems, the accuracy of land use change detection, and the representation of rare land cover transitions.

Discussion

The statistical approach used in the FCPF Carbon Fund project employs a systematic 1km x 1km sampling grid, enabling consistent and reliable long-term monitoring of land cover changes. This design ensures a robust methodology with a relative sampling error of less than 15%, making it suitable for high-accuracy assessments over time.

While effective in maintaining consistency, the approach is limited in its temporal and spatial flexibility. Each new time period requires reanalyzing approximately 4,000 high-resolution imagery sites, which can be resource intensive. Additionally, the method is valid only for the specific area where it was originally implemented, restricting its adaptability to new regions without significant investment.

To apply this approach for assessing a specific value chain, a customized sampling design would be necessary. Such customization would need to account for the unique geographical and operational characteristics of the value chain, ensuring the sampling aligns with its scope and objectives.

Furthermore, achieving higher levels of precision requires a significant increase in the number of samples, with sample size growing exponentially as desired accuracy improves. This presents a practical challenge, particularly for large-scale assessments.

Overall, the statistical approach provides a consistent and reliable framework for long-term monitoring, but its limitations in flexibility, resource demands, and scalability are to be considered carefully when applying it to specific contexts, such as value chain assessments.

Tree Planting

Agroforestry systems, particularly those involving cacao cultivation, and reforestation efforts pose unique challenges for remote sensing due to their diverse and complex nature. These systems often consist of a mix of tree species, crops, and other vegetation, making accurate classification difficult. Lower spatial resolution imagery frequently struggles to capture the small-scale practices or distinguish individual trees within agroforestry and reforestation interventions. Additionally, optical remote sensing methods often lack the capacity to provide detailed information about the vertical structure of vegetation, a critical factor in assessing both agroforestry systems and reforestation interventions.

The assessment of agroforestry systems and reforestation interventions also necessitates the ability to monitor individual growing trees, which can take several years to mature. Identifying these trees in the first two years after planting is particularly challenging with lower spatial resolution imagery. This requires a long series of high-resolution imagery, which can be both costly and complex to analyze. However, recent advancements have introduced datasets capable of providing tree canopy height for specific years or offering very high-resolution imagery. These datasets aim to map the maximum tree height within each 10m pixel, specifically capturing the height of the 98th percentile tree rather than the average height within the pixel. If updated frequently, these datasets hold significant potential for enabling effective monitoring of agroforestry and reforestation efforts, bridging current gaps in data availability and analysis capabilities.

Selected datasets

ETH Global Sentinel-2 10m Canopy Height (2020)

The ETH-GTCH dataset (Lang et al., 2023)provides a global, high-resolution (10m) canopy height map for the year 2020. This dataset addresses the critical need for detailed vegetation height data, which is essential for understanding the global carbon cycle, ecosystem health, and biodiversity conservation. By combining GEDI LiDAR data with dense Sentinel-2 satellite imagery through a probabilistic deep learning model, it delivers accurate and reliable estimates of canopy height. Furthermore, the model incorporates uncertainty management, enhancing its credibility and utility for various environmental monitoring applications.

WRI & Meta Global 1m Tree Canopy Height Map

The WRI/Meta-GTCH dataset (Tolan et al., 2024) represents a groundbreaking advancement in tree monitoring, offering the first-ever global canopy height map at an unprecedented 1-meter resolution. This dataset enables the detection and measurement of individual trees worldwide, providing unparalleled detail. Powered by Meta's DiNOv2 Self-Supervised Learning (SSL) model, the dataset is built from 18 million high-resolution satellite images (0.5m Maxar imagery), collected between 2009 and 2020, with 80% of the data sourced from the period between 2018 and 2020. The model achieves a mean absolute error of 2.8 meters in canopy height prediction, making it a highly accurate resource for applications such as forest management, agroforestry monitoring, and biodiversity conservation.

Method Overview

A quantitative assessment of the datasets for tree planting monitoring was not conducted in this study, as such an evaluation would require extensive and resource-intensive fieldwork beyond the scope of this analysis. Instead, the methodology focused on a comprehensive review of scientific publications associated with these datasets.

To evaluate the usability of each dataset for tree planting and agroforestry monitoring, we conducted a structured assessment of their strengths and weaknesses. This evaluation was based on a systematic review of scientific literature, focusing on three key aspects. First, we examined validation quality, assessing whether the dataset was validated using appropriate methods and if the results demonstrated robust accuracy in detecting tree height and changes over time. Second, we examined temporal consistency, assessing whether the dataset provided reliable and frequent updates for long-term monitoring of tree planting activities. Third, we analyzed spatial resolution and the ability to identify single trees, determining whether the dataset could accurately capture small-scale planting efforts within agroforestry landscapes. Lastly, we reviewed limitations in detecting subtle differences in low canopy heights, which are critical for distinguishing young plantations and assessing early-stage tree growth. By evaluating these criteria, the analysis provided insights into the datasets' suitability for tree planting and agroforestry monitoring, while highlighting specific areas where improvements are needed to enhance their utility in complex agroforestry landscapes.

Overall Assessment

The overall assessment of tree height datasets, as presented in Table 5, revealed notable differences in their performance across key criteria. The ETH-GTCH scored highest with a final score of 73%, driven by its strong performance in usability, completeness (58%), and data management (100%), reflecting robust documentation and accessibility practices. In comparison, the WRI/Meta-GTCH achieved a final score of 50%, performing similarly in usability (66%) but lower in completeness (44%) and data management (24%), indicating gaps in coverage and metadata practices.

Table 5 Overall assessment scores for tree height datasets

	Usability	*Completeness*	*Data Management*	*Final Score*
WRI/Meta-GTCH	Limited	44%	24%	50%
ETH-GTCH	Limited	58%	100%	73%

ETH Global Sentinel-2 10m Canopy Height (2020)

Self-reported Validation

The ETH-GTCH was validated using multiple independent reference datasets and comparison with existing global-scale canopy height estimates.

Comparison with the University of Maryland Canopy Height Map

The ETH-GTCH dataset was compared against the University of Maryland (UMD) Global Canopy Height Map, which combines GEDI data (RH95) with Landsat composites. This comparison showed that the ETH map significantly reduced underestimation bias from −7.1m to −1.7m when validated against GEDI hold-out data (87 million footprints). While the UMD dataset consistently underestimated canopy height, especially for trees taller than 30m, the ETH dataset outperformed UMD across all height ranges, albeit with a tendency to overestimate vegetation heights below 5m and a moderate overestimation bias (≈2m) for trees between 5–20m.

Validation with Independent Airborne LiDAR Data

The ETH-GTCH dataset was further validated against independent airborne LiDAR datasets, including:

- NASA's Land, Vegetation, and Ice Sensor (LVIS) campaigns
- High-resolution canopy height models from small-footprint airborne laser scanning (ALS) campaigns in Europe

In nine out of twelve regions, the ETH dataset demonstrated lower random error (RMSE and MAE) and bias compared to the UMD dataset. While UMD systematically underestimated canopy height across all regions, ETH tended to overestimate tree height in most cases, particularly in high-canopy regions such as Oregon (USA) and Gabon, where UMD had an overall bias ranging between -19% to -23%, while ETH had a bias of -4% to -13% and 7% to 10% respectively.

Discussion

Usability

The ETH-GTCH dataset, utilizing Sentinel-2's 10m resolution, provides a consistent and scalable approach for mapping canopy height globally within a single year.

One of the strengths of this dataset is its enhanced accuracy in estimating tall canopies (>30m), which are key indicators of high carbon stocks. These taller canopy measurements are particularly important for studies focused on climate change and biodiversity, where accurate representation of mature forests can guide decision-making and policy formulation.

A significant limitation is the dataset's tendency to overestimate areas with very low canopy heights (<5m). This overestimation can lead to errors in land cover classifications, particularly in ecosystems with low vegetation such as grasslands or sparse forests and cocoa agroforestry. These inaccuracies may impact the ability to discern between forested areas and other land types, which is crucial for accurate land use, agroforestry, and reforestation monitoring.

The overestimation of low canopy heights is especially problematic in agroforestry landscapes, where trees can be sparse and are interspersed with crops. This issue might lead to misclassification of agroforestry areas and hinder the monitoring of newly planted systems. While the dataset is updated annually, its limitations in detecting low canopies, combined with its 10m spatial resolution, make it unsuitable for accurately identifying trees less than two years of age. Enhancing the dataset's sensitivity to these landscapes would be essential for improving its applicability to agroforestry monitoring.

Completeness

The ETH-GTCH dataset offers strong capabilities in temporal resolution, geographic coverage, validation, and methodological transparency, making it a valuable resource for monitoring tree height dynamics over time. The dataset provides excellent temporal resolution, allowing for frequent updates that support the monitoring of cocoa-driven deforestation. It also has global coverage, making it relevant for assessing cocoa landscapes at a broad scale. Its reliability is further reinforced by rigorous validation protocols and a peer-reviewed publication, ensuring transparency and credibility.

However, the dataset does have some limitations. While the spatial resolution is good, it is less granular than some other datasets, which limits precision at the plot level for cocoa landscapes. The maintenance reliability is poorly documented, meaning it might not be actively updated or improved, which raises concerns about its long-term sustainability as a monitoring tool.

Data management

The ETH-GTCH dataset is hosted on a platform that allows access through an API, enabling seamless integration into analytical workflows. Additionally, it supports data downloads for offline analysis, allowing users to work with the dataset in various environments. The dataset is made available under a Creative Commons Attribution 4.0 International License, ensuring open access and ease of use. Furthermore, its metadata is published on Google Earth Engine, a widely recognized data platform that enhances discoverability.

A key strength of this dataset is its detailed and well-documented metadata, which provides clear and comprehensive information on its methodology, structure, and limitations. This level of transparency improves usability and facilitates informed decision-making when applying the dataset to research and monitoring efforts.

WRI & Meta Global 1m Tree Canopy Height Map

Reported Validation

The WRI/Meta-GTCH was validated through multiple quantitative and qualitative assessments, using airborne LiDAR, GEDI satellite measurements, human-annotated tree detection labels, and comparisons with existing canopy height datasets.

Comparison with Airborne LiDAR Data

The dataset was evaluated against high-resolution aerial LiDAR-derived canopy height models (CHMs) across different test regions, including California (USA) and São Paulo (Brazil). Accuracy metrics included Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and bias (Mean Error - ME). The MAE ranged from 2.6m to 5.2m, with lower errors in regions with shorter vegetation. RMSE values were slightly higher, ranging from 4.4m to 7.5m, reflecting increased uncertainty in taller canopies.

Comparison with Existing Canopy Height Maps

The dataset was compared to previous global canopy height maps, specifically Lang et al. (2022) and Potapov et al. (2021). Visual and statistical comparisons showed that the WRI & Meta dataset provided higher resolution and more detailed canopy height information, particularly for detecting individual trees in sparse and fragmented landscapes.

Validation with GEDI Satellite Data

To assess global consistency, the dataset was validated against GEDI RH95 metrics, which measure the 95th percentile of canopy heights globally. A sample of 20,000 GEDI test points was used, weighted to ensure a proportional representation across different canopy height distributions. The correlation coefficient (R^2) between the WRI & Meta dataset and GEDI RH95 varied from 0.22 (Central

Asia) to 0.73 (North America), with an overall global R^2 of 0.61. A small negative bias was observed, but applying a GEDI calibration model improved the accuracy and reduced bias.

Human-Annotated Tree Detection Validation

In addition to canopy height accuracy, the model's ability to segment tree vs. non-tree pixels was evaluated using a dataset of human-annotated Maxar imagery. Annotators labeled trees over 1m in height with a canopy diameter greater than 3m. The model achieved a user's accuracy (precision) of 82–90% and a producer's accuracy (recall) of 86–88%, demonstrating strong agreement with human-labeled data. The Intersection Over Union (IOU) score ranged from 0.74 to 0.77, reflecting good segmentation accuracy.

Geographic Generalization and Segmentation Metrics

The model's generalizability was tested across multiple geographic regions, showing consistent segmentation accuracy even in areas where training data was limited. Additional segmentation metrics, including Edge Error (EE), measured the sharpness of tree cover boundaries, with an EE score of 0.49–0.52, indicating good boundary definition. The self-supervised learning approach used in training enabled the model to perform well across diverse landscapes.

Discussion

Usability

The WRI/Meta-GTCH dataset demonstrates excellent performance in both dense and sparse tree areas. With its 1-meter resolution, it accurately detects individual trees and significantly outperforms the ETH Global Canopy Height dataset in regions with low tree and open canopies. This level of detail makes it highly valuable for monitoring isolated trees and small clusters, which are often critical components of agroforestry systems and fragmented landscapes.

However, the dataset is limited by its inconsistent temporal coverage. Built from high-resolution imagery collected between 2009 and 2020 with 80% of the data from the period between 2018 and 2020. Despite this concentration, global coverage remains inconsistent, and users cannot rely on the availability of year-specific data for all locations. This irregular temporal coverage poses significant challenges for agroforestry and reforestation monitoring, as up-to-date imagery for specific sites or years may not be available, hindering efforts to track changes or monitor tree growth over time.

Furthermore, the dataset necessitates careful verification of imagery acquisition dates. Analyses using this dataset must account for the exact timing of the imagery to ensure accurate and contextually relevant assessments. Without precise date verification, the risk of temporal mismatches could undermine the reliability of results, particularly in dynamic or rapidly changing environments. This temporal limitation also makes the dataset unsuitable for identifying trees less than two years of age, as detecting recently planted trees requires more frequent updates than the dataset currently provides. Despite these limitations, the WRI & Meta Global Tree Canopy Height Map remains a powerful tool for applications requiring fine-scale tree height data, particularly in sparse tree landscapes.

Completeness

The dataset's spatial resolution is excellent, offering a high level of granularity that allows for precise analysis of cocoa landscapes at the plot level. Additionally, it provides comprehensive global coverage, making it a valuable resource for monitoring the cocoa value chain across different regions. The dataset also benefits from rigorous validation protocols, ensuring reliability, and its methodology has been published in a credible peer-reviewed journal, further enhancing transparency and credibility.

However, the dataset is very limited in temporal resolution and scope, meaning it is not updated frequently enough to support real-time deforestation monitoring and lacks the long-term historical data necessary for GHG assessments. Additionally, its maintenance reliability is poorly documented, raising concerns about its long-term usability as a sustainable monitoring tool. In terms of historical impact, while the dataset has been used in research, there are only limited credible reports of its direct application in shaping policies or driving large-scale interventions.

Data management

The WRI/Meta-GTCH dataset has notable strengths in data accessibility and management, but also key limitations in metadata quality. It is hosted on a platform that enables access via API, facilitating seamless integration into various analytical workflows. Additionally, it allows data downloads and offline analysis, making it a versatile resource for different use cases. The dataset is also provided under a Creative Commons Attribution 4.0 International License, ensuring open access and ease of use. Furthermore, its metadata is published on Google Earth Engine, a well-known data platform, which enhances visibility and discoverability.

However, the main limitation in terms of data management is the quality of the metadata itself. At the time of this report, the metadata lacks critical information about several key elements of the dataset, reducing transparency and making it more challenging for users to fully assess its applicability. Improving the completeness and clarity of the metadata would significantly enhance the dataset's usability and reliability for broader monitoring applications.

Platforms

Selected platforms

Satelligence

Satelligence is a commercial platform that specializes in providing high-accuracy land cover maps tailored for deforestation-free compliance and Scope-3 reporting. With its robust remote sensing capabilities, Satelligence supports organizations in monitoring deforestation and land-use change and ensuring sustainable practices within supply chains.

Global Forest Watch (GFW)

Global Forest Watch (GFW) stands as the most comprehensive public platform for forest monitoring. It provides extensive documentation and exceptional data accessibility, making it a critical resource for researchers, policymakers, and organizations aiming to track and manage forest changes globally. By offering open-access tools and datasets, GFW promotes transparency and empowers stakeholders to engage in evidence-based decision-making for forest conservation.

Mighty Earth

Mighty Earth focuses on mapping the cocoa sourcing locations of major chocolate companies in Côte d'Ivoire and Ghana. Its platform is focused on linking cocoa production to sourcing companies, with the intent to enable stakeholders to address deforestation within the cocoa supply chain.

Trase

Trase provides the most complete cocoa supply chain data for Ghana and Côte d'Ivoire, offering unparalleled insights into the movement of cocoa from production to export. This platform plays a key role in enhancing supply chain transparency and accountability by connecting production regions with buyers and markets. Trase's data-driven approach has been developed to help stakeholders identify and mitigate sustainability risks across the cocoa supply chain.

Method Overview

The assessment of the platforms was conducted using a combination of methods. It primarily involved a detailed review of the documentation provided by each platform, complemented by interviews with key platform managers and technical staff, specifically for Satelligence and Global Forest Watch (GFW). Additionally, the evaluation included an analysis of the methodologies used to combine known datasets, the datasets used, the quality and detail of the platform's documentation, and its overall accessibility to users. While the accuracy of the platforms themselves was not directly measured, their usability was carefully assessed by examining the quality and reliability of the underlying datasets. Particular attention was given to whether the analysis relied solely on global datasets without incorporating locally tailored information, as such an approach may limit the relevance of the results in complex landscapes like those in Côte d'Ivoire and Ghana. The assessment also considered whether the platforms depended on a single dataset rather than employing a more robust convergence of evidence approach, which has been shown to yield more accurate and balanced land cover classifications. Furthermore, we examined whether the platforms used datasets that performed poorly in our accuracy assessment, as reliance on lower-quality inputs could significantly impact the credibility of the platform's outputs. By focusing on these aspects, the evaluation provided a comprehensive understanding of each platform's strengths, limitations, and overall suitability for land cover and supply chain monitoring.

Overall Assessment

The overall assessment of the platforms, presented in Table 6, highlighted varying strengths and weaknesses across key criteria. Global Forest Watch scored the highest with a final score of 63%, driven by good performance in data management (99%) and completeness (86%), although its usability (limited) leaves room for improvement due to its reliance on global dataset, particularly in cocoa landscapes. Satelligence closely followed with a final score of 62%, with a good usability but scoring lower in completeness (43%). Trase achieved a final score of 56%, with balanced performance in usability (good), data management (72%), and inclusiveness (33%), though its completeness (41%) highlights gaps in coverage. Mighty Earth, with a final score of 43%, showed consistent but modest scores across all categories, including usability (very limited), data management (71%), and inclusiveness (33%), reflecting a need for more robust data integration and analysis capabilities.

Table 6 Overall assessment scores for platforms

	Usability	*Completeness*	*Data Management*	*Inclusiveness*	*Final Score*
Satelligence	Good	43%	69%	66%	62%
GFW	Limited	86%	99%	66%	63%
Mighty earth	Very Limited	45%	71%	33%	43%
Trase	Limited	41%	72%	33%	56%

Satelligence

The Satelligence platform delivers high-quality land cover mapping, utilizing robust methodologies and diverse datasets to achieve greater accuracy and completeness compared to publicly available datasets. Its maps are reliable for deforestation-free compliance and sustainability assessments. However, some limitations remain in distinguishing non-forest natural ecosystems and in differentiating between sparse forests and savannah areas, especially in complex landscapes like Ghana. Addressing these challenges could further enhance the platform. Satelligence provides a reliable Scope-3 reporting service, leveraging its locally tailored and accurate land cover maps to deliver detailed insights for carbon accounting and deforestation-free compliance. By integrating a wide range of satellite data sources, including Sentinel-1, Sentinel-2, Landsat, ALOS PALSAR, GEDI, and ICESAT, the platform offers robust aboveground biomass modeling. Occasional validation with airborne LiDAR and field measurements further enhances the reliability of its data. This comprehensive approach positions Satelligence as a valuable resource for Scope-3 reporting. While training materials and metadata are publicly available, some methodological documentation is accessible through Satelligence's platform upon request. However, Satelligence has also published aspects of its methodology on its website in the form of blogs, such as its Sentinel-1 radar change detection method, developed in collaboration with WUR. Additionally, Satelligence presents its work at scientific conferences, including the Living Planet Symposium, where it shares insights on methodologies such as global-scale change detection algorithms for dry and deciduous forests and approaches for an EUDR forest baseline using open data, commodity maps, and forest change detection. These conference presentations are accompanied by publications that further document the methodologies used. Satelligence effectively integrates local data, increasing the platform's relevance compared to global datasets. This integration supports more accurate and context-specific monitoring. However, the platform lacks a formal inclusiveness strategy, which limits its engagement with diverse stakeholders and its ability to ensure data representativeness across all relevant land-use systems. Overall, the Satelligence platform provides a strong and reliable resource for land cover mapping and deforestation-free compliance, with clear areas for further improvement in transparency and stakeholder engagement.

Global Forest Watch

Global Forest Watch (GFW) is recognized as the most complete publicly available platform for forest monitoring. It offers outstanding documentation, comprehensive error handling, and high data completeness, making it highly accessible and reliable for a wide range of users. The platform's GFW Pro version provides advanced features for custom analysis, risk assessment, and tailored monitoring tools, making it particularly suitable for corporate supply chain monitoring and compliance tracking.

GFW primarily utilizes global datasets such as SBTN-NL, GFC Tree Cover, EU-JRC-V2, and GFCF. These datasets are well-documented and hosted on reputable platforms, ensuring transparency and consistency in data analysis. This reliance on standardized global datasets makes GFW highly effective for broad-scale forest monitoring. However, its reliance on global data limits its applicability in more specific contexts, such as agroforestry or cocoa landscapes.

Despite its strengths, GFW's dependence on global datasets results in lower accuracy for monitoring cocoa landscapes in Côte d'Ivoire and Ghana. This highlights the need to integrate locally developed datasets, such as official national land cover data, to improve performance in these regions. Implementing the FAO WHISP Convergence of Evidence approach, which combines results from multiple datasets, is recommended to enhance accuracy and applicability for cocoa landscapes when using GFW.

GFW is actively working to promote inclusivity by supporting projects that assist smallholder farmers in digitizing their cocoa plots and integrating this data into the platform. These efforts aim to enhance local engagement and make monitoring more inclusive, particularly in cocoa-producing regions. Such initiatives highlight GFW's commitment to improving the utility of its platform for local stakeholders.

Overall, Global Forest Watch is an excellent resource for forest monitoring, offering unmatched accessibility and comprehensive tools. However, its effectiveness in specific contexts like cocoa landscapes could be enhanced by incorporating local datasets and applying convergence-based approaches for greater accuracy and inclusivity.

Mighty Earth

Mighty Earth plays a significant role in enhancing data transparency by collaborating with Trase to georeference the sourcing locations of major chocolate companies. This includes mapping cocoa cooperatives and first points of purchase in Côte d'Ivoire and Ghana. By linking cocoa production to sourcing entities, the platform provides valuable insights for understanding supply chain dynamics and their environmental impact.

The platform incorporates global datasets such as Hansen et al. (2013), which defines forests using a 30% tree cover density threshold (CIV – forest precision: 38% ± 9%, recall: 56% ± 13%, GH – forest precision: 44% ± 12%, recall: 46% ± 14%), and the first version of the JRC Forest Map (CIV – forest precision: 48% ± 9%, recall: 89% ± 8%, GH – forest precision: 51% ± 9%, recall: 94% ± 5%).i While these datasets offer broad-scale forest monitoring, their limited accuracy undermines the platform's ability to effectively assess deforestation risks in cocoa landscapes. Mighty Earth utilizes the Abu et al. (2021) dataset for mapping cocoa farms, which has shown limitations in accuracy, with a precision of only 62% and a recall of 83%. This leads to overestimations of cocoa extent and its environmental impact, reducing the reliability of analyses conducted using this dataset. While Mighty Earth is effective as a visualization platform, it lacks tools for comprehensive supply chain analysis or compliance monitoring. Its primary function is to display analysis results rather than to offer actionable tools for assessing supply chain risks or ensuring deforestation-free practices. Although Mighty Earth's use of combined datasets is a positive step, its reliance on low-quality sources limits its overall effectiveness. Incorporating a more diverse range of global and local datasets would enhance the platform's relevance for forest conservation and supply chain accountability.

In summary, while Mighty Earth provides valuable insights into cocoa sourcing locations and deforestation risks, its reliance on limited datasets and lack of robust supply chain analysis tools reduces its overall impact. Greater integration of accurate datasets and the development of supply chain monitoring capabilities would significantly enhance its utility for forest conservation and supply chain accountability.

Trase

Trase provides the most comprehensive data for understanding the cocoa supply chain flows in Ghana and Côte d'Ivoire. The platform offers valuable insights into supply chain dynamics and deforestation risks, helping stakeholders identify potential vulnerabilities and opportunities for improved sustainability practices. Its ability to link cocoa production areas to downstream supply chain actors makes it a powerful tool for analyzing risk at a macro level.

In Côte d'Ivoire, Trase utilizes the ETH satellite-based high-resolution cocoa maps (CIV – cocoa precision: 90% ± 6%, recall: 80% ± 11%, GH – cocoa precision: 78% ± 8%, recall: 86% ± 13%), which demonstrated the highest accuracy for cocoa mapping in this assessment. This robust mapping methodology ensures reliable data on cocoa production areas, enhancing the platform's relevance for understanding cocoa-driven deforestation and supply chain impacts.

Despite its strengths in cocoa mapping, Trase relies on the TMF global forest extent dataset (CIV – forest precision: 50% ± 12%, recall: 33% ± 10%, GH – forest precision: 47% ± 13%, recall: 34% ± 13%) for deforestation analysis in Côte d'Ivoire, which performed poorly in this assessment. This reliance limits the platform's accuracy in evaluating deforestation impacts tied to cocoa production.

Trase's primary focus on the entire value chain, rather than on specific farms or sites, presents limitations for deforestation attribution. This broad approach makes it challenging to provide the detailed, site-specific information required for deforestation-free compliance. As a result, the platform is more suitable for risk analysis than for generating precise compliance monitoring data.

In summary, Trase is a highly effective platform for supply chain analysis and risk assessment, particularly in identifying trends and impacts across the cocoa value chain. However, its reliance on limited forest extent data and its broad focus reduces its effectiveness for site specific compliance monitoring.

Conclusion

Global and local datasets each present distinct advantages and limitations for land cover and carbon monitoring. Global datasets, with their standardized methodologies, transparency, and widespread accessibility, enable consistent comparisons across regions. However, their performance falls short in complex and fragmented landscapes, such as those in Côte d'Ivoire and Ghana, where they struggle to capture nuanced dynamics. In contrast, local datasets, including official national maps, provide higher accuracy and context-specific relevance, making them more effective for monitoring in these regions. With the land cover map 2020, Côte d'Ivoire has made great progress toward open-access data, offering well-documented datasets with robust validation, setting a strong example for public data management. Nonetheless, many local datasets still lack proper documentation, accessibility, and long-term support, limiting their broader applicability. Similarly, public and commercial tools demonstrate a comparable trade-off. Public datasets, such as Global Forest Carbon Fluxes, offer a comprehensive foundation for carbon flux analysis but rely on suboptimal baselines, like Hansen et al., which are insufficient for complex landscapes. Commercial platforms, such as Satelligence, overcome these challenges by delivering more accurate and tailored insights through advanced aboveground biomass modeling, validated with ground and LiDAR data, making them more reliable for Scope-3 reporting and specific use cases.

To date, no single dataset, including commercial ones that generally demonstrate higher levels of accuracy, has achieved consistently high accuracy across all land cover categories. This study suggests that combining datasets through a convergence of evidence approach could offer a more balanced land cover classification by leveraging the strengths of multiple datasets. In this study, we demonstrate an example of such a combination, guided by the results of a rigorous validation process. However, a key limitation is that the same validation data were used both to select the best-performing datasets and to evaluate the final composite map, which may introduce bias and inflate accuracy estimates. Further research is needed to determine the optimal dataset combinations for maximizing recall or precision in specific categories and to robustly validate the results using independent reference data. Nevertheless, the high level of accuracy achieved in this study underscores the potential effectiveness of the convergence of evidence approach, as proposed by FAO WHISP, when paired with a robust dataset selection process. This method provides a structured approach to improving our understanding of the confidence level in land cover classification within complex landscapes by integrating multiple datasets.

In practice, this approach can be applied to:

- Identify land cover change using either the best available dataset or a combination of datasets, assessing how many datasets agree on the same change as an initial confidence indicator.
- Leverage category-level confidence indicators derived from dataset agreement to better define baseline forest extent, reducing false positives that could lead to unnecessary field verification. A more accurate baseline allows for prioritizing field visits in areas that were more likely to have been forests, ensuring that deforestation assessments focus on the most relevant locations.

- Apply a filtering step to discard events that appear highly unlikely after a simple visual check of satellite imagery, improving efficiency in monitoring efforts.

Agroforestry systems, particularly cocoa landscapes, present unique challenges for monitoring due to their inherent complexity, small-scale practices, and low canopy heights. These characteristics make it difficult for conventional remote sensing approaches to deliver accurate and representative results. Current datasets, including high-resolution imagery and tree height maps, offer promising advancements but require further development and integration to effectively address these challenges. Datasets such as the ETH Canopy Height Map and WRI & Meta Tree Canopy Height Map demonstrate significant potential for agroforestry monitoring, providing unprecedented detail at global and local scales. However, limitations like the overestimation of low canopies and inconsistent temporal coverage emphasize the need for improvements and careful dataset selection. Achieving accurate and reliable tree planting monitoring will require combining these emerging tools with robust methodologies tailored to the unique characteristics of agroforestry systems and reforestation activities.

The platforms analyzed in this study each offer unique strengths and address different aspects of land cover and supply chain monitoring, but all have areas for improvement. Satelligence provides accurate land cover and biomass data, making it particularly strong for Scope-3 reporting and local data integration. However, transparency could be enhanced by improving documentation and metadata accessibility and developing a formal inclusiveness strategy to engage a broader range of stakeholders. Global Forest Watch (GFW) is a leading public platform, offering comprehensive tools and datasets for forest monitoring. While GFW is highly accessible and widely used, it struggles with accuracy in cocoa landscapes due to its reliance on global datasets. Incorporating local datasets and leveraging convergence methods could significantly enhance its applicability in these contexts. Mighty Earth plays an important role in mapping cocoa sourcing locations, providing transparency in linking cocoa production to sourcing companies. However, its reliance on low-quality datasets and the absence of robust supply chain analysis tools limits its utility for comprehensive sustainability assessments. Similarly, Trase is highly effective for analyzing supply chains and identifying broad deforestation risks, but its lack of farm-level data limits its ability to support deforestation-free compliance monitoring.

In conclusion, while Global Forest Watch actively seeks to include smallholders through initiatives such as grants for farm digitization in producing countries, other global datasets lack a clear strategy for incorporating local actors into their platforms. This omission is significant, as local knowledge is crucial for refining the context needed to produce highly accurate maps. Furthermore, the exclusion of smallholders from the calibration and development of zero-deforestation compliance monitoring tools increases the risk of further marginalizing these populations.

To mitigate this risk, it is essential to verify deforestation risk through field assessments rather than excluding farmers based solely on satellite imagery. Without proper representation in global datasets, smallholders may be unfairly penalized due to the low accuracy of these products in the regions where they operate. This highlights the critical need for greater inclusivity in the development and calibration of monitoring systems.

Additionally, datasets do not remain valid indefinitely. A structured feedback process that allows users to report inconsistencies (e.g., false positives) and ensures that learnings are integrated into future dataset updates is essential for long-term reliability and improvement. This iterative process would help maintain the accuracy and relevance of monitoring platforms as data evolves, ensuring that they continue to support effective, fair, and inclusive deforestation compliance monitoring.

Recommendations

For WCF and its members

Compliance with Zero Deforestation Regulations and Certification

Selecting a reliable baseline that accurately distinguishes between cocoa, forests, and other tree commodities is critical for assessing deforestation risks in the supply chain. In contrast, global datasets proved inadequate for differentiating between forests and cocoa and are not recommended unless corrected using local data, such as official national datasets. Among available options, Satelligence offers a robust land cover map, which, despite some limitations, demonstrated higher accuracy than publicly available datasets

For any analysis reliant on publicly available datasets, we strongly advocate adopting the convergence of evidence approach, as proposed by FAO WHISP, when paired with a robust dataset selection process to address inherent limitations and biases.

For open public platforms, Global Forest Watch (GFW) and its Pro version are recommended for tailored supply chain analysis. However, these platforms would greatly benefit from the integration of more local data to enhance the accuracy of information currently based on global datasets.

The following recommendations are intended for companies that choose to base their analyses solely on publicly available open-access datasets, instead of using commercial services that may have higher accuracy data available, such as Satelligence. Each scenario outlines the use of remote sensing for compliance with zero-deforestation regulations and certification standards, providing guidance on selecting the most suitable datasets and methodologies to maximize accuracy and reliability within the limitations of public data.

Jurisdictional Risk Assessment for EUDR Compliance

To minimize the risk of excluding forest areas and potentially overlooking deforestation, we recommend prioritizing recall over precision using datasets with moderate precision (>0.7) and high recall (>0.8) that adhere to the FAO forest definition and data specifically developed for 2020. Additionally, to identify areas where cocoa is most likely to have driven deforestation, datasets should meet the same precision and recall thresholds (>0.7 precision, >0.8 recall) for cocoa extent in 2020. Deforestation data should cover the period from 2020 onward and be updated annually to ensure accurate and timely monitoring.

Table 7 Open-access public datasets recommendation for jurisdictional risk assessment for EUDR compliance

CIV	GH
Forest: National data (p: 73 ± 10, r: 81 ± 12) To better assess the confidence level of forest extent data from national datasets, additional lower-quality datasets can be used to explore their level of agreement, such as EU-JRC-V2 (p: 64 ± 12, r: 80 ± 12) and SBTN-NL (p: 57 ± 11, r: 81 ± 11). **Cocoa**: ETH-Cocoa (p: 91 ± 7, r: 80 ± 12) **Deforestation**: GFC	**Forest**: No dataset meets the required precision. Agreement among EU-JRC-V2 (p: 63 ± 11, r: 88 ± 7), national data (p: 63 ± 13, r: 63 ± 13), and SBTN-NL (p: 66 ± 11, r: 87 ± 8) should be used to assess confidence in forest classification. **Cocoa**: ETH-Cocoa (p: 78 ± 8, r: 87 ± 13). **Deforestation**: GFC.

Farm-Level Monitoring for EUDR Compliance

To reduce the need for extensive field verification of farm-level deforestation and minimize the risk of excluding forest areas, which could result in missed deforestation events, we recommend using datasets with both high precision (>0.8) and high recall (>0.8) for forest identification, following the FAO forest definition and utilizing data specifically developed for 2020. Deforestation data should cover the period from 2020 onward and be updated annually to ensure accurate and timely monitoring.

Table 8 Open-access public datasets recommendation for farm-level monitoring for EUDR compliance

CIV	GH
Forest: National data (p: 73 ± 10, r: 81 ± 12). A trade-off on precision reduces the risk of missing deforestation but increases the need for field validation. To better assess the confidence level of forest extent data from national datasets, additional lower-quality datasets can be used to explore their level of agreement, such as EU-JRC-V2 (p: 64 ± 12, r: 80 ± 12) and SBTN-NL (p: 57 ± 11, r: 81 ± 11). **Deforestation**: GFC, TMF	**Forest**: No dataset meets the required precision. Agreement among EU-JRC-V2 (p: 63 ± 11, r: 88 ± 7), national data (p: 63 ± 13, r: 63 ± 13), and SBTN-NL (p: 66 ± 11, r: 87 ± 8) should be used to assess confidence in forest classification. **Deforestation**: GFC.

No farm should be excluded from a supply chain solely based on remote sensing data. Any suspected deforestation must be verified through on-the-ground field validation to ensure accuracy and fairness in decision-making.

Jurisdictional Pre-2020 Deforestation Risk Assessment

To minimize the risk of excluding forest areas and potentially overlooking deforestation, we recommend using datasets with moderate precision (>0.7) and high recall (>0.8) to identify forest extent at the cut-off year. Additionally, to identify areas where cocoa is most likely to have driven deforestation, datasets should meet the same precision and recall thresholds (>0.7 precision, >0.8

recall) for cocoa extent at the cut-off year. Deforestation data should cover the period from the cut-off year onward and be updated annually to ensure accurate and timely monitoring. The following recommendations are intended for companies that choose to base their analyses exclusively on publicly available, open-access datasets. While commercial providers offer services specifically tailored to these tasks, the accuracy of those services, particularly for pre-2020 baseline assessments, was not evaluated in this study.

In both Ghana and Côte d'Ivoire, no publicly available, open-access, high-quality forest extent data is available for pre-2020 dates. Therefore, we recommend dividing the analysis into two distinct periods: one set of datasets should be used to assess changes prior to 2020, while a different selection should be applied from 2020 onward to ensure consistency and accuracy in forest monitoring.

Forest (Pre-2020): In both countries, the recommended approach is to combine multiple publicly available datasets, identifying areas where datasets intersect to improve confidence. Relevant datasets include DW (2015–2025), GFC (2000–2023), GLAD-Forest (2000–2022), GLC-FCS30D (1985–2022), and TMF (1990–2023). For specific cases, such as Rainforest Alliance (2014 cut-off), tailored proprietary datasets developed by the certifier should be preferred when independent validation confirms the reliability and quality of the data. **Forest (2020 onward):** Same recommendation as for EUDR. **Cocoa:** No cocoa maps exist before 2020. If the cut-off year is close to 2020 (up to 2015), ETH-Cocoa (p: 91 ± 7, r: 80 ± 12) could be used as a proxy. **Deforestation:** GFC.

Pre-2020 Zero-Deforestation Compliance Monitoring

To reduce the need for extensive field verification of farm-level deforestation and minimize the risk of excluding forest areas, which could lead to missed deforestation events, we recommend using datasets with both high precision (>0.8) and high recall (>0.8) for forest identification, following the FAO forest definition and utilizing data aligned with the certification's pre-2020 cutoff date. Since no single publicly available, open-access, high-quality dataset exists for forest extent before 2020, the best approach is to combine multiple publicly available datasets to improve confidence. Deforestation data should cover the period up to the cutoff date. The following recommendations are intended for companies that choose to base their analyses exclusively on publicly available, open-access datasets. While commercial providers offer services specifically tailored to these tasks, the accuracy of those services, particularly for pre-2020 baseline assessments, was not evaluated in this study.

Forest (Pre-2020): In both countries, the recommended approach is to combine multiple publicly available datasets, identifying areas where datasets intersect to improve confidence. Relevant datasets include DW (2015–2025), GFC (2000–2023), GLAD-Forest (2000–2022), GLC-FCS30D (1985–2022), and TMF (1990–2023). For specific cases, such as Rainforest Alliance (2014 cut-off), tailored proprietary datasets developed by the certifier should be preferred when independent validation confirms the reliability and quality of the data. **Forest (2020 onward):** Same recommendation as for EUDR. **Deforestation:** GFC, TMF.

No farm should be excluded from a supply chain solely based on remote sensing data. Any suspected deforestation must be verified through on-the-ground field validation to ensure accuracy and fairness in decision-making.

Carbon Analysis and Scope-3 Reporting

The Global Forest Carbon Fluxes dataset is a strong publicly available option for carbon analysis and Scope-3 reporting. However, it relies heavily on Hansen et al. as its forest cover and loss baseline, which has proven suboptimal in complex and fragmented landscapes such as cocoa-growing regions in Côte d'Ivoire and Ghana. Users should exercise caution when utilizing this dataset in these contexts, as it may not fully capture the complex dynamics of these landscapes. Specifically, given the high recall and very low precision identified for GFC in this study, using these datasets is likely to result in a significant overestimation of emissions. For recent years, field verification can help estimate the extent of this overestimation. For older deforestation events, the use of high-resolution imagery, when available, can support the quantification of this overestimation. On the other hand, TMF, which also offers a long time-series of forest loss and degradation, shows higher precision but lower recall. As a result, it is likely to underestimate carbon emissions, since significant areas of forest, and associated loss, may not be captured in the analysis.

To be suitable for Scope-3 reporting, datasets must provide a consistent time series of at least 20 years, yet establishing a reliable forest baseline remains a significant challenge due to limited high-resolution historical imagery. While publicly available datasets such as Hansen et al. (2013), TMF, and GLAD offer long-term deforestation tracking, their limitations in Côte d'Ivoire and Ghana must be carefully considered. A convergence of evidence approach, combining multiple datasets and thoroughly assessing uncertainties, may improve accuracy. However, further research is needed to validate this method for robust long-term deforestation monitoring.

For more tailored assessments, commercial solutions like those offered by Satelligence, which incorporate locally gathered data and advanced modeling techniques, present a robust alternative.

The following recommendations outline the use of remote sensing for both emission and removal assessment.

Deforestation-Related Carbon Emission Assessment

The GFCF dataset is the only public resource for carbon flux analysis. However, its reliance on (Hansen et al., 2013)data makes it less reliable in complex cocoa landscapes like Côte d'Ivoire and Ghana. More information about the use of remote sensing data for this intended use can be found in the GHG accounting manual for cocoa developed by Quantis and WCF (Rizzo et al., 2025).

Carbon Removal Assessment in Cocoa Agroforestry Systems

No public datasets meet the required granularity and local calibration needed for accurate carbon removal assessment. More information about the use of remote sensing data for this intended use, based on drone imagery or LiDAR, can be found in the GHG accounting manual for cocoa developed by Quantis and WCF (Rizzo et al., 2025).

Tree Planting Monitoring

Datasets such as the ETH Canopy Height Map and WRI & Meta Tree Canopy Height Map exhibit considerable potential for tree planting monitoring, offering unprecedented levels of detail at both global and local scales. These datasets provide valuable insights into canopy height and vegetation structure, which are critical for understanding agroforestry systems and reforestation projects. However, their limitations, such as the overestimation of low canopies and inconsistent temporal coverage, highlight the need for further development and careful dataset selection to ensure reliability and accuracy in monitoring.

We strongly recommend investing in additional research to address these challenges and enhance the capabilities of existing datasets. Moreover, the development of solutions specifically tailored to tree planting monitoring, with a focus on addressing the complexity and diversity of such systems, will be essential for advancing sustainable land-use practices, delivery of effective GHG removal projects, and supporting compliance with environmental and sustainability goals.

The following recommendations outline the use of remote sensing for both recently planted agroforestry systems and mature ones.

Monitoring Tree Planting and Agroforestry Establishment (New and Growing Systems)

Monitoring tree planting and agroforestry establishment in new and growing systems requires very high-resolution data to accurately track planted tree status and detect tree mortality. Currently, only drone imagery and field visits provide the granularity and local calibration needed to assess early-stage tree growth with sufficient precision. Publicly available datasets lack the necessary resolution and calibration, making them unsuitable for monitoring agroforestry interventions. To ensure effective tracking, data should be updated annually, integrating on-the-ground verification.

Monitoring Plot-Level Agroforestry and Tree Cover Stability (Mature Systems)

Monitoring plot-level agroforestry and tree cover stability in mature systems requires high-resolution tree canopy height data to effectively differentiate between cocoa and shade trees, with updates on an annual basis. While ETH-GTCH and UMD-GFCH offer sufficient temporal coverage, their coarse resolution and lack of precision for small canopies limit their reliability in agroforestry contexts. In contrast, WRI/Meta-GTCH provides the necessary spatial resolution, but its irregular temporal coverage makes it unreliable for consistently tracking tree cover changes across different sites and years. To ensure accurate and continuous monitoring, data sources must offer both high spatial precision and regular updates, which are currently not fully met by publicly available datasets.

For Platforms

Satelligence

Enhance Transparency: Publish documentation and metadata in recognized public repositories, such as Dataverse, to improve transparency and ensure long-term availability for a wide range of stakeholders. This will make the platform's methodologies and data more accessible, fostering trust and collaboration.

Inclusiveness Strategy: Develop a structured inclusiveness strategy to engage a broader range of stakeholders, particularly smallholders and local communities. While efforts are already being made, such as collaborations with Fairtrade, where Satelligence's data is used to verify smallholder farms, and those farms, in turn, contribute to improving commodity maps and baseline accuracy, a more systematic approach would help ensure that inclusivity is consistently applied across all aspects of data collection and analysis. Additionally, better communication of these existing efforts would increase transparency and demonstrate the platform's commitment to inclusive and representative data practices, ultimately enhancing the accuracy and acceptance of its outputs.

Global Forest Watch

Integrate High Quality Datasets: While Global Forest Watch (GFW) is highly accessible and widely used, its reliance on global datasets limits its accuracy in complex landscapes such as cocoa-growing regions. Incorporating local or proprietary datasets of proven great quality through independent validation, including official national land cover data, would significantly improve its precision and relevance in these contexts.

Mighty Earth

Enhance Dataset Integration: Integrate additional global and local datasets, including those aligned with the FAO WHISP Convergence of Evidence approach, to improve accuracy and applicability for cocoa-producing regions. Incorporating locally developed datasets, such as official national land cover data, would further enhance the platform's relevance and precision.

Adopt Robust Cocoa Map: Utilize more robust datasets, such as the ETH cocoa dataset, or incorporate official national cocoa data to achieve more accurate and reliable assessments of cocoa production areas. This will help address current limitations in mapping cocoa extents and their environmental impacts.

Trase

Enhance Deforestation Assessments: To improve the accuracy of deforestation assessments, Trase should incorporate multiple forest extent datasets, particularly those aligned with the FAO WHISP Convergence of Evidence approach. This methodology would allow Trase to leverage the strengths of various datasets, reducing biases and improving reliability.

Integrate Local Data: Include locally developed datasets, such as official national forest cover data, to provide more accurate and region-specific insights. This would enhance the platform's relevance and applicability in complex landscapes like cocoa-growing regions in Côte d'Ivoire and Ghana.

For Countries

Follow Côte d'Ivoire's Example: Countries should take inspiration from the 2020 land cover map of Côte d'Ivoire, which sets a benchmark for thorough documentation, robust validation, and publication in global repositories. This approach ensures transparency, reliability, and accessibility for a wide range of stakeholders, making it a model for improving national land cover datasets.

Improve Documentation and Accessibility: Many countries still require significant improvements in documentation, data management, metadata, and accessibility. To address these shortcomings, partnerships with established platforms such as Global Forest Watch (GFW) are recommended to enhance data distribution and ensure proper documentation practices. By improving accessibility and adopting best practices in data management, locally tailored datasets could become more integrated and sustainable, supporting more effective land-use monitoring and compliance efforts.

Develop High-Quality Historical Data: While some official national historical land cover datasets exist, they often lack the consistency and robustness of newer datasets, limiting their reliability for long-term analysis. Countries are strongly encouraged to invest in the development of high-quality historical datasets using consistent methodologies to ensure comparability over time. These efforts would significantly enhance long-term carbon assessments and strengthen national monitoring systems, ensuring that datasets remain reliable, actionable, and aligned with international standards.

References

Angelica Rizzo, Alexi Ernstoff, Alexandra Stern, Thierry Hohmann, Sophie Chalumeau, Paula Sanginés de Cárcer, Michael Matarasso, & Elizabeth Howard. (2025). GHG ACCOUNTING MANUAL FOR COCOA A step by step GHG Accounting methodology for Land Use Change, Land Management and Removals Prepared by. *Wolrd Cocoa Foundation*.

BNETD. (2024). *Carte d'occupation des sols de Côte d'Ivoire en 2020*. https://africageoportal.maps.arcgis.com/apps/webappviewer/index.html?id=88c2493e722546c09c2a0a8b394c4454

Bourgoin, C., Verhegghen, A., Carboni, S., Ameztoy, I., Ceccherini, G., Colditz, R., & Achard, F. (2024). *Global map of forest types 2020 - version 0*. European Commission, Joint Research Centre (JRC).

Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., & Tait, A. M. (2022). Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data 2022 9:1*, *9*(1), 1–17. https://doi.org/10.1038/s41597-022-01307-4

FC. (2020). *Ghana National Land Use/ Land Cover Map Interactive Map*. https://ghana-national-landuse.knust.ourecosystem.com/interface/

Forest Data Partnership. (2024). *Cocoa model 2024a*. https://github.com/google/forest-data-partnership/tree/main/models/cocoa

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/SUPPL_FILE/HANSEN.SM.PDF

Harris, N. L., Gibbs, D. A., Baccini, A., Birdsey, R. A., de Bruin, S., Farina, M., Fatoyinbo, L., Hansen, M. C., Herold, M., Houghton, R. A., Potapov, P. V., Suarez, D. R., Roman-Cuesta, R. M., Saatchi, S. S., Slay, C. M., Turubanova, S. A., & Tyukavina, A. (2021). Global maps of twenty-first century forest carbon fluxes. *Nature Climate Change 2021 11:3*, *11*(3), 234–240. https://doi.org/10.1038/s41558-020-00976-6

Kalischek, N., Lang, N., Renier, C., Daudt, R. C., Addoah, T., Thompson, W., Blaser-Hart, W. J., Garrett, R., Schindler, K., & Wegner, J. D. (2023). Cocoa plantations are associated with deforestation in Côte d'Ivoire and Ghana. *Nature Food 2023 4:5*, *4*(5), 384–393. https://doi.org/10.1038/s43016-023-00751-8

Lang, N., Jetz, W., Schindler, K., & Wegner, J. D. (2023). A high-resolution canopy height model of the Earth. *Nature Ecology & Evolution 2023 7:11*, *7*(11), 1778–1789. https://doi.org/10.1038/s41559-023-02206-6

Liu, L., Zhang, X., & Zhao, T. (2023). *GLC_FCS30D: the first global 30-m land-cover dynamic monitoring product with fine classification system from 1985 to 2022*. https://doi.org/10.5281/ZENODO.8239305

Mazur, E., Sims, M., Goldman, E., Schneider, M., Daldoss Pirri, M., Beatty, C. R., Stolle, F., & Stevenson, M. (2025). *SBTN Natural Lands Map-Technical Documentation Version 1.1*.

Parente, L., Sloat, L., Mesquita, V., Consoli, D., Stanimirova, R., Hengl, T., Bonannella, C., Teles, N., Wheeler, I., Hunter, M., Ehrmann, S., Ferreira, L., Mattos, A. P., Oliveira, B., Meyer, C., Şahin, M., Witjes, M., Fritz, S., Malek, Ž., & Stolle, F. (2024). *Global Pasture Watch - Annual grassland class and extent maps at 30-m spatial resolution (2000—2022)*. https://doi.org/10.5281/ZENODO.13890401

Potapov, P., Hansen, M. C., Pickens, A., Hernandez-Serna, A., Tyukavina, A., Turubanova, S., Zalles, V., Li, X., Khan, A., Stolle, F., Harris, N., Song, X. P., Baggett, A., Kommareddy, I., & Kommareddy, A. (2022). The Global 2000-2020 Land Cover and Land Use Change Dataset Derived From the Landsat Archive: First Results. *Frontiers in Remote Sensing*, *3*, 856903. https://doi.org/10.3389/FRSEN.2022.856903/BIBTEX

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C. E., Armston, J., Dubayah, R., Blair, J. B., & Hofton, M. (2021). Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, *253*, 112165. https://doi.org/10.1016/J.RSE.2020.112165

Satelligence. (2024). *Satelligence - Sustainability monitoring simplified*. https://satelligence.com/

Tolan, J., Yang, H. I., Nosarzewski, B., Couairon, G., Vo, H. V., Brandt, J., Spore, J., Majumdar, S., Haziza, D., Vamaraju, J., Moutakanni, T., Bojanowski, P., Johns, T., White, B., Tiecke, T., & Couprie, C. (2024). Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar. *Remote Sensing of Environment*, *300*, 113888. https://doi.org/10.1016/J.RSE.2023.113888

Vancutsem, C., Achard, F., Pekel, J. F., Vieilledent, G., Carboni, S., Simonetti, D., Gallego, J., Aragão, L. E. O. C., & Nasi, R. (2021). Long-term (1990–2019) monitoring of forest cover changes in the humid tropics. *Science Advances*, *7*(10). https://doi.org/10.1126/SCIADV.ABE1603/SUPPL_FILE/ABE1603_SM.PDF

WRI, & WBCSD. (2022). *Land Sector and Removals Guidance Pilot Testing and Review Draft Part 2*.

Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., Cartus, O., Santoro, M., Fritz, S., Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, L., Tsendbazar, N.-E., … Arino, O. (2021). *ESA WorldCover 10 m 2020 v100*. https://doi.org/10.5281/ZENODO.5571936

Annexes

Annex 1

Annex 1 provides an overview of the selected datasets, including links to their data repositories and references to the corresponding scientific publications.

Land cover datasets

Title	Abbreviation	Source	Reference
Satelligence	**Satelligence**	Closed dataset https://satelligence.com/	(Satelligence, 2024)
Carte Occupation du Sol Ivoirien 2020	**Official-CIV**	https://africageoportal.maps.arcgis.com/home/item.html?id=46348aa12a3d46e592584737de64f72a	(BNETD, 2024)
Tthe national land use map for Ghana	**Official-GH**	https://ghana-national-landuse.knust.ourecosystem.com/interface/	(FC, 2020)
Cocoa Probability model 2024a and Forest Persistence v0	**FDaP**	https://developers.google.com/earth-engine/datasets/publisher/forestdatapartnership	(Forest Data Partnership, 2024)
SBTN Natural Lands Map 2020 v1.1	**SBTN-NL**	https://developers.google.com/earth-engine/datasets/catalog/WRI_SBTN_naturalLands_v1#description	(Mazur et al., 2025)
Dynamic World V1	**DW**	https://developers.google.com/earth-engine/datasets/catalog/GOOGLE_DYNAMICWORLD_V1	(Brown et al., 2022)
Global forest cover 2020 V1	**JRC-EU-V1**	https://developers.google.com/earth-engine/datasets/catalog/JRC_GFC2020_subtypes_V0	(Bourgoin et al., 2024)
Global forest cover 2020 V2	**JRC-EU-V2**	https://forobs.jrc.ec.europa.eu/GFC	(Bourgoin et al., 2024)
Tropical Moist Forests product – Annual Change v1 2020	**TMF**	https://forobs.jrc.ec.europa.eu/TMF	(Vancutsem et al., 2021)

ESA World Cover	**ESA-WOC**	https://esa-worldcover.org/en/data-access	(Zanaga et al., 2021)
Hansen Global Forest Change 10%	**GFC-10%**	https://storage.googleapis.com/earthenginepartners-hansen/GFC-2023-v1.11/download.html	(Hansen et al., 2013)
Hansen Global Forest Change 30%	**GFC-30%**	https://storage.googleapis.com/earthenginepartners-hansen/GFC-2023-v1.11/download.html	(Hansen et al., 2013)
Forest extent, 2020	**GLAD-Forest**	https://glad.umd.edu/dataset/GLCLUC2020	(Potapov et al., 2022)
ETH high-resolution maps of cocoa	**ETH-Cocoa**	https://www.research-collection.ethz.ch/handle/20.500.11850/654400	(Kalischek et al., 2023)
Global 30-meter Land Cover Change Dataset	**GLC_FCS30D**	https://gee-community-catalog.org/projects/glc_fcs/	(Liu et al., 2023)
Global Pasture Watch	**GPW**	https://zenodo.org/records/11281157	(Parente et al., 2024)

Carbon/biomass datasets

Title	Abbreviation	Source	Reference
Satelligence	**Satelligence**	Closed dataset https://satelligence.com/	(Satelligence, 2024)
Global Forest Carbon Fluxes	**GFCF**	https://gee-community-catalog.org/projects/cflux/	(Harris et al., 2021)

Tree canopy height

Title	Abbreviation	Source	Reference
WRI & Meta Global 1m Tree Canopy Height Map	**WRI/Meta-GTCH**	https://gee-community-catalog.org/projects/meta_trees/#dataset-citation	(Tolan et al., 2024)
ETH Global Sentinel-2 10m Canopy Height (2020)	**ETH-GTCH**	https://www.research-collection.ethz.ch/handle/20.500.11850/609802	(Lang et al., 2023)
Global Forest Canopy Height, 2019	**UMD-GFCH**	https://glad.umd.edu/dataset/gedi	(Potapov et al., 2021)

Annex 2

Annex 2 provides an overview of the characteristics of the selected datasets, including available land cover categories (F: Forest, C: Cocoa, NF: Non-forest natural lands), spatial resolution, time period, and the pre-processing steps applied for the analysis.

Land cover datasets

	Available categories	Resolution	Temporal coverage	Pre-processing	Licence	Update
Satelligence	F,C,NF	10m	2020	No processing	Closed	no clear schedule
Official-CIV	F,C,NF	10m	2020	**Forest:** dense forest, light forest, forest gallery, secondary forest/degraded forest, mangrove, forest plantation/reforestation, swamp forest/forest on hydromorphic soil, tree savannah **Cacao:** cocoa plantation **Other natural**: shrub formations/ thickets, herbaceous formations **Other:** all the other categories	Creative Commons Attribution 4.0 International	no clear schedule
Official-GH	F,C,NF	10m	2019 & 2021(draft)	**Forest:** close Forest, open forest, mangrove **Cocoa:** mono cocoa, shaded cocoa **Other natural:** grassland **Other:** all other categories	Unknown	no clear schedule
FDaP	C	10m	2020	**Cocoa**: cocoa probability model 2024a > 0.5 **Forest**: not cocoa and forest Persistence v0 >0.95 **Other**: all other pixels	Creative Commons Attribution Non Commercial 4.0 International Commercial access can be granted upon demand	no clear schedule

SBTN-NL	F,NF	30m	2020	**Forest**: natural forests, mangroves, wet natural forests, natural peat forests **Other natural**: natural short vegetation, wet natural short vegetation, natural peat short vegetation **Other**: all the other categories	Creative Commons Attribution Share Alike 4.0 International	no clear schedule
DW	TC,NF	10m	2015 - 2025	Each pixel is assigned a class based on the highest probability and reclassified following the following rules **Forest**: Trees **Other natural**: grass, shrub and scrub	Creative Commons Attribution 4.0 International	2 weeks
JRC-EU-V1	F	10m	2020	No processing	All data are provided free-of charge, without restriction of use. For the full license information see the Copernicus Regulation of the European Commission.	no clear schedule
JRC-EU-V2	F	10m	2020	No processing	All data are provided free-of charge, without restriction of use. For the full license information see the Copernicus Regulation of the European Commission.	no clear schedule
TMF	F	30m	1990 - 2023	**Forest**: undisturbed tropical moist forest, degraded tropical moist forest, tropical moist forest regrowth		annually

ESA-WOC	TC	10m	2020 - 2021	**Forest**: tree cover, mangroves **Other Natural**: shrubland, grassland, moss and lichen	Creative Commons Attribution 4.0 International	no clear schedule
GFC-10%	TC	30m	2000 – 2023	**Forest**: union of more than 10% canopy cover and no loss observed before 2020 and tree gain **Other**: all other pixels	Creative Commons Attribution 4.0 International	annually
GFC-30%	TC	30m	2000 - 2023	**Forest**: union of more than 30% canopy cover and no loss observed before 2020 and tree gain **Other**: all other pixels	Creative Commons Attribution 4.0 International	annually
GLAD-Forest	F	30m	2000 & 2020	No processing	Creative Commons Attribution License	no clear schedule
ETH-Cocoa	C	10m	2020	No processing, used official cocoa/other dataset (0.65 threshold)	Creative Commons Attribution 4.0 International	no clear schedule
GLC_FCS30D	F,NF	30m	1985-2000 (every 5 years), 2000-2022 annual	**Forest**: all forest type and mangrove **Other Natural**: all shrubland, all sparse vegetation, swamp, marsh **Other**: all other class	Creative Commons Attribution 4.0 International	no clear schedule
GPW	NF	30m	2000-2020	**Other Natural**: threshold: 0.42 on natural/semi-natural dataset based on documentation.	Creative Commons Attribution 4.0 International	Update announced with no clear timeline, likely annually

5

Carbon/biomass datasets

	Resolution	Temporal coverage	License	Update
Satelligence	10m	2001-2024	Closed	annually
GFCF	30m	2001-2023	Creative Commons Attribution 4.0 International License	annually

Tree canopy height

	Resolution	Temporal coverage	License	Update
WRI/Meta-GTCH	1m	composite of image between 2009-2020	Creative Commons Attribution 4.0 International License	no clear schedule
ETH-GTCH	10m	2020	Creative Commons Attribution 4.0 International License	no clear schedule
UMD--GFCH	30m	2019	No license	Update announced without timeline

Annex 3

Annex 3 consists of an Excel file that provides a comprehensive overview of the quality assessment results for all evaluated datasets and platforms. This supplementary material contains detailed scoring and evaluation metrics across key assessment categories, including accuracy, completeness, data management, and inclusiveness.

The Excel file is structured into multiple sheets, each corresponding to different dataset types (land cover, tree height, biomass/carbon monitoring) and platforms. It presents quantitative scores alongside explanatory notes to clarify specific strengths and limitations identified during the assessment. The final ranking of datasets and platforms is calculated using the weighted methodology described in the main report, ensuring transparency in how the results were derived.

This annex serves as a detailed reference for users seeking in-depth insights into the performance of individual datasets and platforms, supporting informed decision-making in dataset selection and use.

Annex 4

Annex 4 consists of a detailed table summarizing the results of the independent accuracy assessment conducted on 19 different land cover maps, organized in alphabetical order for ease of reference. This annex provides a systematic evaluation of the accuracy of each dataset, ensuring a transparent comparison of their performance in key land cover categories.

The table includes the following metrics for each assessed dataset:

- **Overall Accuracy:** The percentage of correctly classified land cover across all categories.
- **Precision:** The likelihood that a given land cover classification is correct.
- **Recall:** The ability of the dataset to capture the full extent of a specific land cover type.
- **Class-Specific Performance:** Accuracy breakdowns for key categories such as cocoa, forest, non-forest natural lands, and other land types.
- **Confidence Intervals:** Statistical uncertainty ranges for precision and recall values to indicate the robustness of the assessment.

dataset	country	category	precision	ci precision	recall	ci recall
Composite	civ	other	71.99	11.42	81.01	9.55
		other natural	48.83	35.65	32.93	27.56

		forest	84.24	11.29	71.35	12.22
		cocoa	79.94	8.61	87.69	10.99
		overall accuracy	75.21	6.97		
	gh	other	78.78	12.17	39.42	11.18
		other natural	42.12	9.22	89.69	8.26
		forest	78.57	14.63	53.73	13.64
		cocoa	73.36	10.64	87.09	13.49
		overall accuracy	59.62	6.75		
DW	civ	other	100	0	13.53	8.44
		other natural	36.04	26.27	45.68	29.12
		forest	35.63	6.55	97.4	2.92
		overall accuracy	41.05	7.24		
	gh	other	90.41	8.81	31.99	8.88
		other natural	47.4	11.82	71.86	12.96
		forest	42.66	8.41	82.56	10.46
		overall accuracy	53.25	6.68		
ESA-WOC	civ	other	88.89	20.74	13.6	8.48
		other natural	29.29	15.29	79.87	21.75
		forest	35.78	7.68	79.55	9.59
		overall accuracy	39.11	7.53		
	gh	other	95.27	6.41	33.83	9.06
		other natural	43	9.38	91.82	7.69
		forest	41.28	10.12	56.09	13.86
		overall accuracy	52.42	6.68		
ETH-Cocoa	civ	other	95.89	2.85	98.21	1.33
		cocoa	90.63	6.98	80.43	11.8
		overall accuracy	95.07	2.62		
	gh	other	98.42	1.8	97.26	1.05
		cocoa	78.5	8.44	86.5	13.41
		overall accuracy	96.15	1.86		
EU-JRC-V1	civ	other	93.17	5.16	61.35	9.69

		forest	48.47	9.16	88.99	8.02
		overall accuracy	69.36	7.39		
	gh	other	97.66	2.32	72.21	5.77
		forest	51.49	8.97	94.46	5.43
		overall accuracy	77.51	4.86		
EU-JRC-V2	civ	other	90.69	6.34	81.45	8.06
		forest	63.74	11.56	79.59	11.92
		overall accuracy	80.91	6.7		
	gh	other	95.44	2.85	82.47	5.75
		forest	62.54	11.19	88.13	7.35
		overall accuracy	83.88	4.83		
GLAD-Forest	civ	other	94.38	7.38	38.2	9.96
		forest	38.97	7.01	94.55	7.01
		overall accuracy	54.79	7.67		
	gh	other	97.59	2.66	72.53	5.44
		forest	51.75	8.63	94.26	6.27
		overall accuracy	77.7	4.63		
GLAD-LC	civ	other	99.5	1.02	17.45	9.16
		other natural	29	19.34	58.11	28.32
		forest	38.97	7.01	94.38	6.99
		overall accuracy	43.74	7.4		
	gh	other	95.94	5.29	41.16	9.18
		other natural	42.43	10.83	66.8	12.36
		forest	53.16	8.73	91.55	7.86
		overall accuracy	59.19	6.51		
GLC-FCS30D	civ	other	91.87	10.69	12.03	7.71
		other natural	41.35	42.08	23.98	28.05
		forest	32.44	6.19	95.03	4.04
		overall accuracy	37.69	7.06		
	gh	other	99.85	0.31	18.52	7.79
		other natural	36.64	13.16	41.45	14.36

		forest	33.42	6.84	89.1	8.21
		overall accuracy	40.87	6.61		
GPW	civ	other	92.58	4.53	99.62	0.65
		other natural	23.4	39.07	1.45	2.15
		overall accuracy	92.25	4.55		
	gh	other	82.55	4.82	96.65	2.96
		other natural	53.94	28.82	16.12	11.05
		overall accuracy	80.88	4.99		
FDaP	civ	other	67.54	7.11	94.67	5.35
		forest	60.62	15.92	33.99	11.93
		cocoa	96.41	4.25	44.26	12.24
		overall accuracy	68.72	6.32		
	gh	other	74.32	5.18	98.04	2.1
		forest	67.79	17.77	24.37	12.08
		cocoa	82.1	14.58	25.03	9.59
		overall accuracy	74.04	4.91		
GFC-10%	civ	other	92.44	10.49	26	9.24
		forest	35.11	6.38	94.96	6.95
		overall accuracy	46.45	7.52		
	gh	other	99.36	0.89	41.51	7.42
		forest	36.04	6.94	99.2	1.12
		overall accuracy	55.89	6.47		
GFC-30%	civ	other	77.64	7.77	62.56	8.75
		forest	38.05	9.09	56.07	12.58
		overall accuracy	60.68	7.35		
	gh	other	84.07	4.95	82.83	4.14
		forest	43.61	11.64	45.83	13.94
		overall accuracy	74.52	4.86		
National data	civ	other	77.12	8.21	84.02	7.25
		other natural	100	0	0.72	1.48
		forest	73.33	9.85	80.73	11.82
		cocoa	72.24	13.98	41.09	11.79

		overall accuracy	75.4	5.96		
	gh	other	85.97	4.96	89.4	5.07
		other natural	41.3	9.8	77.35	11.74
		forest	63.42	12.6	62.51	13.34
		cocoa	83.02	12.24	64.08	11.81
		overall accuracy	80.64	4.81		
SBTN-NL	civ	other	84.66	8.63	68.99	10.19
		other natural	36.29	27.83	33.52	28.04
		forest	57.22	11.12	81.41	11.28
		overall accuracy	69.51	7.6		
	gh	other	82.82	6.28	74.96	8.36
		other natural	49.01	15.73	43.25	14.09
		forest	66.08	11.3	86.76	8.2
		overall accuracy	70.95	6.19		
Satelligence	civ	other	83.31	9.78	86.41	4.02
		other natural	58	13.82	19.99	12.05
		forest	87.41	6.13	70.48	12.61
		cocoa	75.42	8.93	90.26	10.61
		overall accuracy	82.53	5.89		
	gh	other	93.61	5.05	79.09	4.49
		other natural	60	13.72	63.4	11.92
		forest	59.33	11.29	82.59	9.84
		cocoa	78.04	8.41	87.4	13.43
		overall accuracy	80.76	4.71		
TMF	civ	other	76.82	6.81	86.74	5.08
		forest	49.61	12.08	32.51	10.07
		overall accuracy	71.21	6.45		
	gh	other	82.13	5.89	70.68	6.71
		forest	47.48	12.6	34.11	12.96
		overall accuracy	62.54	6.41		

Annex 5

Annex 5 consists of a detailed table summarizing the results of the independent accuracy assessment conducted on 19 different land cover maps, grouped by key categories (cocoa, forests and other naturasl lands) and organized in alphabetical order for ease of reference.

The table includes the following metrics for each assessed dataset:

- **Precision:** The likelihood that a given land cover classification is correct.
- **Recall:** The ability of the dataset to capture the full extent of a specific land cover type.
- **Confidence Intervals:** Statistical uncertainty ranges for precision and recall values to indicate the robustness of the assessment.

Cocoa

	CIV		GH	
Dataset	**precision**	**recall**	**precision**	**recall**
Composite	80 ± 9	88 ± 11	73 ± 11	87 ± 13
ETH-Cocoa	91 ± 7	80 ± 12	78 ± 8	87 ± 13
FDaP	96 ± 4	44 ± 12	82 ± 15	25 ± 10
National data	72 ± 14	41 ± 12	83 ± 12	64 ± 12
Satelligence	75 ± 9	90 ± 11	78 ± 8	87 ± 13

Forest

	CIV		GH	
Dataset	**precision**	**recall**	**precision**	**recall**
Composite	84 ± 11	71 ± 12	79 ± 15	54 ± 14
DW	36 ± 7	97 ± 3	43 ± 8	83 ± 10
ESA-WOC	36 ± 8	80 ± 10	41 ± 10	56 ± 14
EU-JRC-V1	48 ± 9	89 ± 8	51 ± 9	94 ± 5
EU-JRC-V2	64 ± 12	80 ± 12	63 ± 11	88 ± 7
GLAD-Forest	39 ± 7	95 ± 7	52 ± 9	94 ± 6
GLAD-LC	39 ± 7	94 ± 7	53 ± 9	92 ± 8
GLC_FCS30D	32 ± 6	95 ± 4	33 ± 7	89 ± 8
FDaP	61 ± 16	34 ± 12	68 ± 18	24 ± 12
GFC-10%	35 ± 6	95 ± 7	36 ± 7	99 ± 1
GFC-30%	38 ± 9	56 ± 13	44 ± 12	46 ± 14
National data	73 ± 10	81 ± 12	63 ± 13	63 ± 13
SBTN-NL	57 ± 11	81 ± 11	66 ± 11	87 ± 8
Satelligence	87 ± 6	70 ± 13	59 ± 11	83 ± 10
TMF	50 ± 12	33 ± 10	47 ± 13	34 ± 13

Other natural lands

	CIV		GH	
Dataset	**precision**	**recall**	**precision**	**recall**
Composite	49 ± 36	33 ± 28	42 ± 9	90 ± 8
DW	36 ± 26	46 ± 29	47 ± 12	72 ± 13
ESA-WOC	29 ± 15	80 ± 22	43 ± 9	92 ± 8
GLAD-LC	29 ± 19	58 ± 28	42 ± 11	67 ± 12
GLC_FCS30D	41 ± 42	24 ± 28	37 ± 13	41 ± 14
GPW	23 ± 39	1 ± 2	54 ± 29	16 ± 11
National data	100 ± 0	1 ± 1	41 ± 10	77 ± 12
SBTN-NL	36 ± 28	34 ± 28	49 ± 16	43 ± 14
Satelligence	58 ± 14	20 ± 12	60 ± 14	63 ± 12