ReforesTree: A Dataset for Estimating Tropical Forest Carbon Stock with Deep Learning and Aerial Imagery

Gyri Reiersen[1,2], David Dao[2], Björn Lütjens[3], Konstantin Klemmer[4,5], Kenza Amara[2], Attila Steinegger[6], Ce Zhang[2], Xiaoxiang Zhu[1]

[1] Technical University of Munich
[2] ETH Zurich
[3] Massachusetts Institute of Technology
[4] University of Warwick
[5] New York University
[6] WWF Switzerland
gyri.reiersen@tum.de, david.dao@inf.ethz.ch

Abstract

Forest biomass is a key influence for future climate, and the world urgently needs highly scalable financing schemes, such as carbon offsetting certifications, to protect and restore forests. Current manual forest carbon stock inventory methods of measuring single trees by hand are time, labour, and cost intensive and have been shown to be subjective. They can lead to substantial overestimation of the carbon stock and ultimately distrust in forest financing. The potential for impact and scale of leveraging advancements in machine learning and remote sensing technologies is promising, but needs to be of high quality in order to replace the current forest stock protocols for certifications.

In this paper, we present **ReforesTree**, a benchmark dataset of forest carbon stock in six agro-forestry carbon offsetting sites in Ecuador. Furthermore, we show that a deep learning-based end-to-end model using individual tree detection from low cost RGB-only drone imagery is accurately estimating forest carbon stock within official carbon offsetting certification standards. Additionally, our baseline CNN model outperforms state-of-the-art satellite-based forest biomass and carbon stock estimates for this type of small-scale, tropical agro-forestry sites. We present this dataset to encourage machine learning research in this area to increase accountability and transparency of monitoring, verification and reporting (MVR) in carbon offsetting projects, as well as scaling global reforestation financing through accurate remote sensing.

Introduction

The degradation of the natural world is unprecedented in human history and a key driver of the climate crisis and the Holocene extinction (Ceballos and Ehrlich 2018). Forests play a significant role in the planet's carbon cycle, directly impacting local and global climate through its biogeophysical effects and as carbon sinks, sequestering and storing carbon through photosynthesis (Griscom et al. 2017).

However, since the year 2000, we have lost 361 million ha of forest cover, equivalent to the size of Europe, mainly in tropical areas (Hansen et al. 2013). This accounts for 18%

of global anthropogenic emissions and contributes to driving up atmospherical carbon levels (IPCC 2019). Forests, especially tropical forests, also provide habitats for 80% of land-based biodiversity and with the increasing risk and frequency of wildfires, droughts, and extreme weather, forest ecosystems are under severe pressure (Shi et al. 2021).

To avoid planetary tipping points (Rockstöm et al. 2009) and maintain a stable and livable climate, mankind urgently need to reduce carbon emissions until 2050 and restore essential ecosystems (IPCC 2021). Forests and natural carbon sequestration are important climate change mitigation strategies (Canadell and Raupach 2008) with a biophysical mitigation potential of 5,380 MtCO2 per year on average until 2050 (IPCC 2019).

Forestry is a large industry and the causes of deforestation are mostly economically driven (FAO 2020) (Geist and Lambin 2001). For the last 20 years, major conservation efforts have been underway to mitigate and safeguard against these losses. One of the global financing strategies is carbon offsets (Blaufelder et al. 2021). Initially, it started as the Clean Development Mechanism (CDM) under the Kyoto Protocol, allowing governments and business organizations from industrialized countries to invest in forestry in developing countries by buying carbon credits to offset industrialized emissions (FAO 2020) Several other independent bodies have later developed official standards for verifying and certifying carbon offsetting projects, such as the Gold Standard (GS) and the Verified Carbon Standard (VERRA). The certification process for forest carbon offsetting projects is capital and labour intensive, especially due to the high cost of manual monitoring, verification and reporting (MVR) of the forest carbon stock.

The carbon offsetting market is rapidly increasing and expected to grow by a factor of 100 until 2050 due to high demand and available capital (Blaufelder et al. 2021). However, the main obstacle is limited supply of offsetting projects as forest owners lack upfront capital and market access (Kreibich and Hermwille 2021).

Recent research investigations (Badgley et al. 2021; West et al. 2020) have shown that the current manual forest carbon stock practices systematically overestimate forestry carbon

offsetting projects with up to 29% of the offsets analyzed, totaling up to 30 million tCO2e (CO_2 equivalents) and worth approximately \$410 million. The overestimation was identified to come from subjective estimations and modeling of the carbon stock baseline and of the project's additionally and leakage reporting. There is thus a need for higher quality carbon offsetting protocols and higher transparency and accountability of the MVR of these projects (Haya et al. 2020).

There are three key aspects that are important for the use of remote sensing in MVR of forest carbon stock. One aspect is financial; using available and accessible technology and sensors to lower the cost and upfront capital requirements for forest owners to get certified, especially in low and middle-income countries. The second aspect is reducing subjectivity in estimating carbon stock and increasing trustworthiness and transparency in the carbon offsetting certification protocols. And lastly, the solutions need to be scalable due to the urgency of financing forest restoration, especially in tropical regions.

Various verification bodies, new ventures, and academia are currently developing remote sensing technologies to automate parts of the certification process of forestry carbon offsetting projects (Narine, Popescu, and Malambo 2020; Dao et al. 2019). Satellite imagery is increasing in quality and availability and, combined with state-of-the-art deep learning and lidar, promises to soon map every tree on earth (Hanan and Anchang 2020) and to enable forest aboveground biomass and carbon to be estimated at scale (Saatchi et al. 2011; Santoro et al. 2021). Compared to current manual estimates, these advancements reduce time and cost and increase transparency and accountability, thus lowering the threshold for forest owners and buyers to enter the market (Lütjens, Liebenwein, and Kramer 2019). Nevertheless, these algorithms risk additionally contributing to the systematic overestimation of carbon stocks, not reducing it, and are not applicable for small-scale forests, below 10,000 ha (White et al. 2018), (Global Forest Watch 2019).

Accurately estimating forest carbon stock, especially for small scale carbon offset projects, presents several interesting machine learning challenges, such as high variance of species and occlusion of individual tree crowns. There are many promising approaches, such as hyperspectral species classification (Schiefer et al. 2020), lidar-based height measurements (Ganz, Käber, and Adler 2019) and individual tree crown segmentation across sites (Weinstein et al. 2020b). However, these applications have been developed mainly on datasets from temperate forests and, to the knowledge of the authors, there is no publicly available dataset of tropical forests with both aerial imagery and ground truth field measurements.

Here, we present **ReforesTree**, a dataset of six tropical agroforestry reforestation project sites with individual tree crown bounding boxes of over 4,600 trees matched with their respective diameter at breast height (DBH), species, species group, aboveground biomass (AGB), and carbon stock. This dataset represents ground truth field data mapped with low-cost, high-resolution RGB drone imagery to be used to train new models for carbon offsetting protocols and for benchmark existing models.
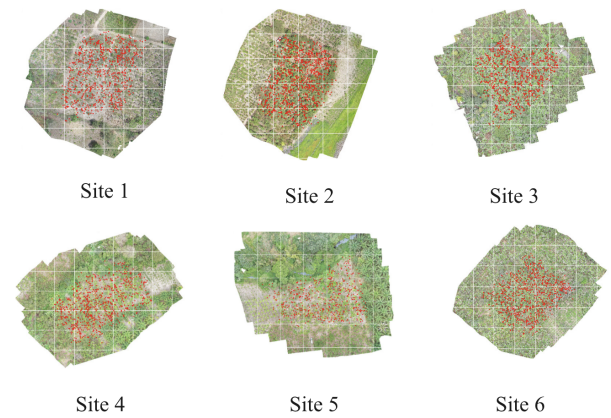

Figure 1: Drone imagery of each site of the ReforesTree dataset with a resolution of 2cm/px. The red dots are the locations of the trees measured in field surveys, plotted to make clear that the coverage of drone images were larger than the field measured area.

To summarize, with **ReforestTree**, we contribute the following: 1) the first publicly available dataset of tropical agro-forestry containing both ground truth field data matched with high resolution RGB drone imagery at the individual tree level and 2) a methodology for reducing the current overestimation of forest carbon stock through deep learning and aerial imagery for carbon offsetting projects.

Related Work

Deep Learning for Remote Sensing

In recent years, deep learning (DL), and especially deep convolutional neural networks (CNN) are increasing in popularity for image analysis in the remote-sensing community (Ma et al. 2019), (Zhu et al. 2017). With the increase in computation power, larger datasets, transfer learning, and breakthroughs in network architecture, DL models have outperformed conventional image processing methods in several image tasks such as land use and land cover (LULC) classification, segmentation and detection. Examples of deep supervised learning in remote sensing are the prediction of wildfires (Yang, Lupascu, and Meel 2021), detection of invasive species (Bjorck et al. 2021). CNNs offer feature extraction capabilities in recognizing patterns in both spatial and temporal data, even with low resolution inputs. With recent advances in meta and few shot learning these models can be trained and generalized on larger datasets and fine-tuned for local variance.

Manual Forest Inventory

The standardized forest carbon stock inventory consists of manually measuring and registering sample trees of a project site. Tree metrics such as diameter at breast height (DBH), height, and species are then put through scientifically developed regression models called allometric equations to cal-
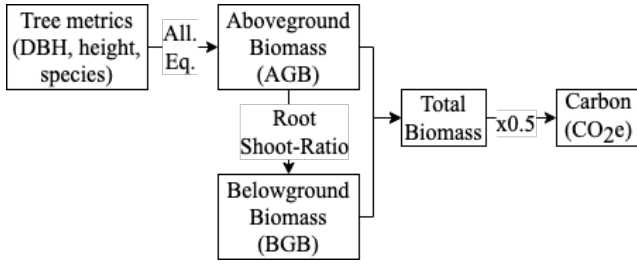
Figure 2: The standard procedure for calculating the correct amount of carbon offsets to be certified for a reforestation project. The tree metrics are collected from manual forest inventory.

SITE NO.	NO. OF TREES	NO. OF SPECIES	SITE AREA	TOTAL AGB	TOTAL CO2E
1	743	18	0.51	8	5
2	929	22	0.62	15	9
3	789	20	0.48	10	6
4	484	12	0.47	5	3
5	872	14	0.56	15	9
6	846	16	0.53	12	7
TOTAL	4463	28	3.17	66	40

Table 1: Overview of the six project sites in Ecuador, as gathered in field measurements. Aboveground biomass (AGB) is measured in metric tons and area in hectares.

culate the aboveground biomass (AGB) as seen in Figure 2. The total biomass of a forest is the total AGB added with the below-ground biomass (BGB), calculated using a root-to-shoot ratio specific to the forest type and region (Ma et al. 2021).

The procedure how to calculate the correct amount of carbon offsets (CO_2e) to be certified for a project is standardized through (Pearson, Walker, and Brown 2005) as shown in Figure 2. The (CO_2e), also known as the baseline forest carbon stock, is equivalent of the total biomass divided by two. Despite being prone to error propagation (Petrokofsky et al. 2012; Malhi et al. 2004) and shown to systematically overestimate carbon stock (Badgley et al. 2021), this is currently the standardized forest inventory method for certification of forestry projects.

Related Methods and Models

The following are three types of methods to estimate forest carbon stock remotely, adapted from (Sun and Liu 2019); 1) inventory-based models, based on national and regional forest inventories and regression models, are known to overestimate due to over-representations of dense commercial forests in the data, (Global Forest Watch 2019). 2) Satellite-based models leveraging datasets from optical remote sensing, synthetic aperture radar satellites (SAR), and lidar (LiDAR) to create global aboveground biomass and carbon maps (Santoro et al. 2021; Saatchi et al. 2011; Spawn, Sullivan, and Lark 2020). 3) Ecosystem-based models using topography, elevation, slope, aspect, and other environmental factors to construct statistical models and quantitatively describe the process of forest carbon cycle to estimate forest carbon stock(Ma et al. 2021).

The most scalable and affordable of these methods are, evidently, satellites-based models. Nevertheless, these models and global maps are yet to estimate carbon stock at local scale and provide accurate estimates of highly heterogeneous and dense forest areas due to their low resolution of 30-300m (Bagheri, Shataee, and Erfanifard 2021). An individual tree-based model that takes the individual overstory trees into account can provide this accuracy, especially if fused with geostatistical and satellite data.

In recent years, researchers have achieved high accuracy

for standard forestry inventory tasks such as individual tree crown detection (Weinstein et al. 2019), lidar-based height estimation (Ganz, Käber, and Adler 2019), and species classification (Miyoshi et al. 2020; Schiefer et al. 2020; Mäyrä et al. 2021), using deep learning models and aerial imagery. This shows high potential for combining high-resolution imagery with deep learning models as a method for accurate carbon stock estimation for small-scale reforestation projects (Sun and Liu 2019).

As most tropical forests are situated in low to middle income countries, without access to hyperspectral, lidar and other more advanced sensors, the models need to be developed using available technologies. A trade-off for accuracy and data availability is basic high-resolution RGB drone imagery. Drone imagery (1-3cm/px resolution), combined with CNN, has previously been used to directly estimate biomass and carbon stock in individual mangrove trees (Jones et al. 2020) or indirectly by detecting species or tree metrics such as DBH or height (Nåfält 2018; Omasa et al. 2003), achieving an accuracy similar to manual field measurements. And by leveraging multi-fusion approaches (Du and Zare 2020; Zhang 2010), e.g. combining low-resolution satellite, high-resolution drone imagery, and field measurements and contextual ecological or topological data, and multi-task learning (Crawshaw 2020), e.g. tree metrics and carbon storage factors as auxiliary tasks, these models can replace and scale the existing manual forest inventory.

There are several datasets for tree detection and classification from drone imagery such as the NEON dataset (Weinstein et al. 2020a), or the Swedish Forest Agency mainly from temperate forests from the US or Europe. To our knowledge, there are no publicly available datasets including both field measurements and drone imagery of heterogeneous tropical forests.

Dataset and Method

The **ReforesTree** dataset consists of six agro-forestry sites in the central coastal region of Ecuador. The sites are of dry tropical forest type and eligible for carbon offsetting certification with forest inventory done and drone imagery captured in 2020. See Table 1 for information on each site.

Forest Inventory Data and Drone Imagery

Field measurements were done by hand for all live trees and bushes within the site boundaries and include GPS location, species, and diameter at breast height (DBH) per tree. Drone imagery was captured in 2020 by an RGB camera from a Mavic 2 Pro drone with a resolution of 2cm per pixel. Each site is around 0.5 ha, mainly containing banana trees (Musaceae) and cacao plants (Cacao), planted in 2016-2019.

$$AGB_{fruit} = 0.1466 * DBH^{2.223} \qquad (1)$$

$$AGB_{musacea} = 0.030 * DBH^{2.13} \qquad (2)$$

$$AGB_{cacao} = 0.1208 * DBH^{1.98} \qquad (3)$$

$$AGB_{timber} = 21.3 - 6.95 * DBH + 0.74 * DBH^2 \qquad (4)$$

The aboveground biomass (AGB) is calculated using published allometric equations for tropical agro-forestry, namely Eq.1 for fruit trees, including citrus fruits (Segura, Kanninen, and Suárez 2006), Eq.2 banana trees (Van Noordwijk et al. 2002), Eq.3 for cacao (Yuliasmara, Wibawa, and Prawoto 2009), and Eq.4 for shade trees (timber) (Brown and Iverson 1992). These are commonly used in global certification standards. The carbon stock is calculated through the standard forest inventory methodology using a root-to-shoot ratio of 22%, which is standard for dry tropical reforestation sites (Ma et al. 2019).

Data Processing and Method

The raw data is processed in several steps as seen in Figure 3. The goal of this process is to have a machine learning ready dataset that consists of matched drone image of an individual tree with the trees labels, such as AGB value. All the drone images have been cropped to fit tightly the boundaries of the field measured areas. The details of this cropping process, and the code repository, are in the Appendix.

Initially the RGB orthomosaics are cut into 4000×4000 tiles and sent through DeepForest, a python package for predicting individual tree crowns from RGB imagery (Weinstein et al. 2019), fine-tuned on some manually labelled bounding boxes from the sites. Afterwards, the bounding boxes containing more than 80% white were filtered out, e.g. bounding boxes lying on the border of the drone imagery, and manually labeled to banana and non-banana, due to the easily recognizable characteristics of banana trees, resulting in clear bounding boxes of all trees as shown in Figure 4.

To fuse the tree information extracted from the ground measurements with the bounding boxes of the trees detected, we used OneForest, a recent machine learning approach for fusing citizen data with drone imagery. To remove noise introduced in both GPS locations, OneForest uses a greedy optimal transport algorithm. This is a known coupling method to map between two GPS positions (center of bounding box from drone imagery and GPS location of tree from field data). Developed by Villani (Villani 2003), the methods finds the minimal distance between two distributions via a convex linear program optimizing for a matching that moves the mass from one distribution to the other with minimal cost. The cost is usually defined as the euclidean distance or the Kulback-Leibler divergence between
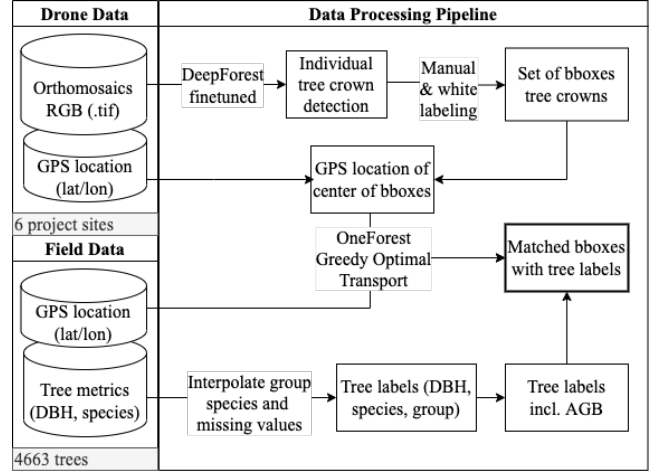

Figure 3: The raw data and data processing pipeline for the ReforesTree dataset, resulting in labels matched to bounding boxes per tree.
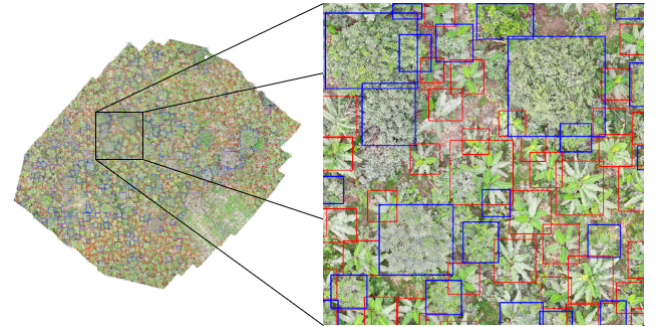

Figure 4: Bounding box annotations per tree, as a result of fine-tuned DeepForest tree crown detection and manual cleaning. Red boxes represent banana trees and blue boxes represent other species.

the distributions. The optimum, i.e. the minimal distance between the two distributions, is called the Wasserstein metric.

Baseline CNN Model

With a dataset of matched bounding boxes and tree labels, we fine-tuned a basic pre-trained CNN, ResNet18 (He et al. 2015) with a mean-square-error loss to estimate individual tree AGB. The results were satisfying despite the simple baseline model, and proves that the individual tree estimation from drone imagery has potential.

Fourteen images were identified as being larger than the expected crown size of a tree, and they were center cropped at 800×800. To preserve the crown size information, the smaller images were zero-padded up to 800×800, before all images were resized to fit the network architecture.

The dataset has is unbalanced with regards to species, of which 43% is cacao and 32% is banana. Additionally, due to the trees being planted between 2016-2019, many of the trees have similar size (e.g. DBH) and half of the trees have DBH between 7-10cm. The training dataset consisted of equal number of samples of species and DBH, and from the different project sites.

Experiments

With the emerging new biomass maps and forest stock estimation models, we used the **ReforesTree** dataset to benchmark these maps and compare with our baseline CNN model for AGB estimation. We compared the maps taken from (Global Forest Watch 2019), (Spawn, Sullivan, and Lark 2020), and (Santoro et al. 2021). The Global Forest Watch's Above-Ground Woody Biomass dataset is a global map of AGB and carbon density at 30m×30m resolution for the year 2000. It is based on more than 700,000 quality-filtered Geoscience Laser Altimeter System (GLAS) lidar observations using machine learning models based on allometric equations for the different regions and vegetation types. The second dataset from (Spawn, Sullivan, and Lark 2020) is a 300m×300m harmonized map based on overlayed input maps. The input maps were allocated in proportion to the relative spatial extent of each vegetation type using ancillary maps of tree cover and landcover, and a rule-based decision schema. The last, and most recent 100m×100m dataset from (Santoro et al. 2021) is obtained by spaceborne SAR (ALOS PALSAR, Envisat ASAR), optical (Landsat-7), lidar (ICE-SAT) and auxiliary datasets with multiple estimation procedures with a set of biomass expansion and conversion factors following approaches to extend ground estimates of wood density and stem-to-total biomass expansion factors.

As seen in Table 2, all of the available global AGB maps have a tendency to overestimate the ground truth measurements up to a factor of ten. These are not encouraging results showing that these maps are far from being accurate enough to be used in remote sensing of forest carbon stock at a small scale, as is the case for the **ReforesTree** dataset.

Our baseline model, on the other hand, has a slight tendency of underestimating the biomass. The model has an evident advantage, to be trained on the dataset, but these initial results show promise for the individual tree estimation approach using drone imagery for forest carbon inventory.

Site no.	Field Data	GFW 2019	Spawn 2020	Santoro 2021	Baseline (Ours)
1	8	99	97	36	7
2	15	108	130	42	8
3	10	36	206	15	15
4	5	5	102	32	9
5	15	73	352	12	11
6	12	26	91	72	15
Tot.	66	331	413	89	65

Table 2: The benchmark results from comparing different models for estimating AGB with the forest inventory of the ReforesTree sites. All numbers are given as AGB in Mg. GFW is (Global Forest Watch 2019), Spawn is (Spawn, Sullivan, and Lark 2020), Santoro is (Santoro et al. 2021). All of these three are satellite-based. Lastly, the baseline CNN is our drone-based model.

Conclusions and Future Work

We introduce the **ReforesTree** dataset in hopes of encouraging the fellow machine learning community to take on the challenge of developing low-cost, scalable, trustworthy and accurate solutions for monitoring, verification and reporting of tropical reforestation inventory. We also present an outlined methodology for creating an annotated machine learning dataset from field data and drone imagery, and train a baseline CNN model for individual tree aboveground biomass estimation. This methodology includes a data processing pipeline leveraging a fine-tuned tree crown detection algorithm and an optimal transport matching algorithm for reduction of GPS noise.

The **ReforesTree** dataset of field measurements and low-cost, high-resolution RGB drone imagery represents the trade-off for accuracy and data availability of remote sensing of forest carbon stock in tropical regions. It can be used to train new or benchmark existing models for MVR of carbon offsetting reforestation protocols. Remote inventory of small scale tropical reforestation projects comes with several ecological challenges, such high biodiversity, level of canopy closure, and topology. This dataset is a start to develop a generalized model that can be fine-tuned on local scale. Future work will investigate ways to improve the methodology and reduce error in the machine learning ready dataset, and increase the explainability to have a trustworthy and transparent model. Additionally, we see further potential in fusing satellite and other available geoecological data layers as well as leveraging the multiple labels available (e.g. DBH, species) as auxiliary tasks in a multitask learning problem.

As the world is rapidly approaching planetary doom, we need to collaborate across disciplines to implement and scale the climate mitigation strategies available. Restoration of forests is one of our most important climate mitigation strategies. And by reducing the overestimation of carbon offsets, we can allow every man on earth who owns a tree to participate in climate action. Biodiverse and sustainable forestry can provide hope not only the for the machine learning community, but also beyond.

Acknowledgments

References

Badgley, G.; Freeman, J.; Hamman, J. J.; Haya, B.; Trugman, A. T.; Anderegg, W. R.; and Cullenward, D. 2021. Systematic over-crediting in California's forest carbon offsets program. *bioRxiv*.

Bagheri, R.; Shataee, S.; and Erfanifard, S. Y. a. 2021. Canopy Based Aboveground Biomass and Carbon Stock Estimation of Wild Pistachio Trees in Arid Woodlands Using GeoEye-1 Images. *Journal of Agricultural Science and Technology*, 23(1).

Barenne, V.; Bohl, J. P.; Dekas, D.; and Engelmann, T. 2022. Tropical Forest Carbon Stock Estimation using RGB Drone Imagery.

Bjorck, J.; Rappazzo, B. H.; Shi, Q.; Brown-Lima, C.; Dean, J.; Fuller, A.; and Gomes, C. 2021. Accelerating Ecological Sciences from Above: Spatial Contrastive Learning for Remote Sensing. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(17): 14711–14720.

Blaufelder, C.; Levy, C.; Mannion, P.; Pinner, D.; and Weterings, J. 2021. McKinsey&Co: A Blueprint for Scaling Voluntary Carbon Markets to Meet the Climate Challenge. Accessed 31.05.2021.

Brown, S.; and Iverson, L. 1992. Biomass estimates for tropical forest. *World Res. Rev.*, 4: 366–383.

Canadell, J. G.; and Raupach, M. R. 2008. Managing Forests for Climate Change Mitigation. *Science*, 320: 1456–1457.

Ceballos, G.; and Ehrlich, P. 2018. The misunderstood sixth mass extinction. *Science*, 360: 1080.2–1081.

Crawshaw, M. 2020. Multi-Task Learning with Deep Neural Networks: A Survey.

Dao, D.; Cang, C.; Fung, C.; Zhang, M.; Pawlowski, N.; Gonzales, R.; Beglinger, N.; and Zhang, C. 2019. GainForest: Scaling Climate Finance for Forest Conservation using Interpretable Machine Learning on Satellite Imagery. *ICML Climate Change AI workshop 2019*.

Du, X.; and Zare, A. 2020. Multiresolution Multimodal Sensor Fusion for Remote Sensing Data With Label Uncertainty. *IEEE Transactions on Geoscience and Remote Sensing*, 58.

FAO. 2020. *Global Forest Resources Assessment 2020: Main report*. FAO. ISBN 978-92-5-132974-0.

Ganz, S.; Käber, Y.; and Adler, P. 2019. Measuring Tree Height with Remote Sensing—A Comparison of Photogrammetric and LiDAR Data with Different Field Measurements. *Forests*, 10: 694.

Geist, H. J.; and Lambin, E. F. 2001. *What drives tropical deforestation?: a meta-analysis of proximate and underlying causes of deforestation based on subnational case study evidence*. LUCC International Project Office, University of Louvain.

Global Forest Watch. 2019. Aboveground Live Woody Biomass Density. Dataset Accessed: 30.11.2021.

Griscom, B. W.; Adams, J.; Ellis, P. W.; Houghton, R. A.; Lomax, G.; Miteva, D. A.; Schlesinger, W. H.; Shoch, D.; Siikamäki, J. V.; Smith, P.; Woodbury, P.; Zganjar, C.; Blackman, A.; Campari, J.; Conant, R. T.; Delgado, C.; Elias, P.; Gopalakrishna, T.; Hamsik, M. R.; Herrero, M.; Kiesecker, J.; Landis, E.; Laestadius, L.; Leavitt, S. M.; Minnemeyer, S.; Polasky, S.; Potapov, P.; Putz, F. E.; Sanderman, J.; Silvius, M.; Wollenberg, E.; and Fargione, J. 2017. Natural climate solutions. *Proceedings of the National Academy of Sciences*, 114(44): 11645–11650.

Hanan, N. P.; and Anchang, J. Y. 2020. Satellites could soon map every tree on Earth. *Nature*, 587.

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.; and Townshend, J. R. G. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, 342(6160): 850–853.

Haya, B.; Cullenward, D.; Strong, A. L.; Grubert, E.; Heilmayr, R.; Sivas, D. A.; and Wara, M. 2020. Managing uncertainty in carbon offsets: insights from California's standardized approach. *Climate Policy*, 20(9): 1112–1126.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2015. Deep Residual Learning for Image Recognition. arXiv:1512.03385.

IPCC. 2019. 2019: Summary for Policymakers. In Shukla, P.; Skea, J.; Buendia, E. C.; Masson-Delmotte, V.; Pörtner, H.-O.; Roberts, D. C.; Zhai, P.; Slade, R.; Connors, S.; van Diemen, R.; Ferrat, M.; Haughey, E.; Luz, S.; Neogi, S.; Pathak, M.; Petzold, J.; Pereira, J. P.; Vyas, P.; Huntley, E.; Kissick, K.; Belkacemi, M.; and Malley, J., eds., *Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems*, 7–11.

IPCC. 2021. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

Jones, A. R.; Raja Segaran, R.; Clarke, K. D.; Waycott, M.; Goh, W. S. H.; and Gillanders, B. M. 2020. Estimating Mangrove Tree Biomass and Carbon Content: A Comparison of Forest Inventory Techniques and Drone Imagery. *Frontiers in Marine Science*, 6: 784.

Kreibich, N.; and Hermwille, L. 2021. Caught in between: credibility and feasibility of the voluntary carbon market post-2020. *Climate Policy*, 21(7): 939–957.

Lütjens, B.; Liebenwein, L.; and Kramer, K. 2019. Machine Learning-based Estimation of Forest Carbon Stocks to increase Transparency of Forest Preservation Efforts. *2019 NeurIPS Workshop on Tackling Climate Change with Machine Learning*.

Ma, H.; Mo, L.; Thomas W. Crowther, D. S. M.; van den Hoogen, J.; Stocker, B. D.; Terrer, C.; and Zohner, C. M.

2021. The global distribution and environmental drivers of aboveground versus belowground plant biomass. *Nature Ecology & Evolution*, 5: 1110–1122.

Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; and Johnson, B. A. 2019. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152: 166–177.

Malhi, Y.; Phillips, O. L.; Chave, J.; Condit, R.; Aguilar, S.; Hernandez, A.; Lao, S.; and Perez, R. 2004. Error propagation and scaling for tropical forest biomass estimates. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1443): 409–420.

Miyoshi, G. T.; Arruda, M. d. S.; Osco, L. P.; Marcato Junior, J.; Gonçalves, D. N.; Imai, N. N.; Tommaselli, A. M. G.; Honkavaara, E.; and Gonçalves, W. N. 2020. A Novel Deep Learning Method to Identify Single Tree Species in UAV-Based Hyperspectral Images. *Remote Sensing*, 12(8).

Mäyrä, J.; Keski-Saari, S.; Kivinen, S.; Tanhuanpää, T.; Hurskainen, P.; Kullberg, P.; Poikolainen, L.; Viinikka, A.; Tuominen, S.; Kumpula, T.; and Vihervaara, P. 2021. Tree species classification from airborne hyperspectral and Li-DAR data using 3D convolutional neural networks. *Remote Sensing of Environment*, 256: 112322.

Narine, L. L.; Popescu, S. C.; and Malambo, L. 2020. Using ICESat-2 to Estimate and Map Forest Aboveground Biomass: A First Example. *Remote Sensing*, 12(11).

Nåfält, S. 2018. Estimating above ground biomass in a Salix plantation using high resolution UAV images. Student thesis series INES, Lund University:8963727.

Omasa, K.; Qiu, G. Y.; Watanuki, K.; Yoshimi, K.; and Akiyama, Y. 2003. Accurate Estimation of Forest Carbon Stocks by 3-D Remote Sensing of Individual Trees. *Environmental Science & Technology*, 37.

Pearson, T.; Walker, S.; and Brown, S. 2005. Sourcebook for BioCarbon Fund Projects. Accessed 15.09.2021 URL: https://winrock.org/document/sourcebook-for-land-use-land-use-change-and-forestry-projects/.

Petrokofsky, G.; Kanamaru, H.; Achard, F.; Goetz, S. J.; Joosten, H.; Holmgren, P.; Lehtonen, A.; Menton, M. C. S.; Pullin, A. S.; and Wattenbach, M. 2012. Comparison of methods for measuring and assessing carbon stocks and carbon stock changes in terrestrial carbon pools. How do the accuracy and precision of current methods compare? A systematic review protocol. *Environmental Evidence*, 1: 6.

Rockstöm, J.; Steffen, W.; K. Noone, Á. P.; Chapin, F. S.; Lambin, E.; Lenton, T. M.; Scheffer, M.; Folke, C.; Schellnhuber, H.; Nykvist, B.; Wit, C. A. D.; Hughes, T.; S. van der Leeuw, H. R.; Sörlin, S.; Snyder, P. K.; R. Costanza, U. S.; Falkenmark, M.; Karlberg, L.; Corell, R. W.; Fabry, V. J.; Hansen, J.; Walker, B.; Liverman, D.; Richardson, K.; Crutzen, P.; and Foley, J. 2009. Planetary boundaries:exploring the safe operating space for humanity. *Ecology and Society*, 14: 32.

Saatchi, S. S.; Harris, N. L.; Brown, S.; Lefsky, M.; Mitchard, E. T. A.; Salas, W.; Zutta, B. R.; Buermann, W.; Lewis, S. L.; Hagen, S.; Petrova, S.; White, L.; Silman, M.;

and Morel, A. 2011. Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, 108(24): 9899–9904.

Santoro, M. 2018. GlobBiomass - global datasets of forest biomass.

Santoro, M.; Cartus, O.; Carvalhais, N.; Rozendaal, D. M. A.; Avitabile, V.; Araza, A.; de Bruin, S.; Herold, M.; Quegan, S.; Rodríguez-Veiga, P.; Balzter, H.; Carreiras, J.; Schepaschenko, D.; Korets, M.; Shimada, M.; Itoh, T.; Moreno Martínez, A.; Cavlovic, J.; Cazzolla Gatti, R.; da Conceição Bispo, P.; Dewnath, N.; Labrière, N.; Liang, J.; Lindsell, J.; Mitchard, E. T. A.; Morel, A.; Pacheco Pascagaza, A. M.; Ryan, C. M.; Slik, F.; Vaglio Laurin, G.; Verbeeck, H.; Wijaya, A.; and Willcock, S. 2021. The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations. *Earth System Science Data*, 13(8): 3927–3950.

Schiefer, F.; Kattenborn, T.; Frick, A.; Frey, J.; Schall, P.; Koch, B.; and Schmidtlein, S. 2020. Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 170: 205–215.

Segura, M.; Kanninen, M.; and Suárez, D. 2006. Allometric models for estimating aboveground biomass of shade trees and coffee bushes grown together. *Agroforestry Systems*, 68: 143–150.

Shi, H.; Tian, H.; Lange, S.; Yang, J.; Pan, S.; Fu, B.; and Reyer, C. P. O. 2021. Terrestrial biodiversity threatened by increasing global aridity velocity under high-level warming. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 18: 36.

Spawn, S.; Sullivan, C.; and Lark, T. e. a. 2020. Harmonized globadgleyl maps of above and belowground biomass carbon density in the year 2010. *Sci Data*, 7: 112.

Sun, W.; and Liu, X. 2019. Review on carbon storage estimation of forest ecosystem and applications in China. *Forest Ecosystems*, 7: 4.

Van Noordwijk, M.; Rahayu, S.; Hairiah, K.; Wulan, Y.; Farida, A.; and Verbist, B. 2002. Carbon stock assessment for a forest-to-coffee conversion landscape in Sumber-Jaya (Lampung, Indonesia): from allometric equations to land use change analysis. *Science in China*, 45.

Villani, C. 2003. *Topics in optimal transportation*. 58. American Mathematical Soc.

Weinstein, B. G.; Marconi, S.; Bohlman, S.; Zare, A.; Singh, A.; Graves, S. J.; and White, E. 2020a. NEON Crowns: a remote sensing derived dataset of 100 million individual tree crowns. *bioRxiv*.

Weinstein, B. G.; Marconi, S.; Bohlman, S.; Zare, A.; and White, E. 2019. Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. *Remote Sensing*, 11(11): 1309.

Weinstein, B. G.; Marconi, S.; Bohlman, S. A.; Zare, A.; and White, E. P. 2020b. Cross-site learning in deep learning RGB tree crown detection. *Ecological Informatics*, 56: 101061.

West, T. A. P.; Börner, J.; Sills, E. O.; and Kontoleon, A. 2020. Overstated carbon emission reductions from voluntary REDD+ projects in the Brazilian Amazon. *Proceedings of the National Academy of Sciences*, 117(39): 24188–24194.

White, A. E.; Lutz, D. A.; Howarth, R. B.; and Soto, J. R. 2018. Small-scale forestry and carbon offset markets: An empirical study of Vermont Current Use forest landowner willingness to accept carbon credit programs. *PLOS ONE*, 13(8): 1–24.

Yang, S.; Lupascu, M.; and Meel, K. S. 2021. Predicting Forest Fire Using Remote Sensing Data And Machine Learning. arXiv:2101.01975.

Yuliasmara, F.; Wibawa, A.; and Prawoto, A. 2009. Carbon stock in different ages and plantation system of cocoa: allometric approach. *Pelita Perkebunan (a Coffee and Cocoa Research Journal)*, 26.

Zhang, J. 2010. Multi-source remote sensing data fusion: status and trends. *International Journal of Image and Data Fusion*, 1.

Zhu, X. X.; Tuia, D.; Mou, L.; Xia, G.-S.; Zhang, L.; Xu, F.; and Fraundorfer, F. 2017. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4): 8–36.

Technical Appendix

Raw data cleaning

All 28 species were divided into 6 species family groups: banana, cacao, fruit, timber, citrus and other.

The field data was manually collected as a standard manual forest inventory, potentially leading to human errors, missing values and outliers.

The dataset needed to reflect the ground truth. Therefore it was important not to remove trees from the dataset unnecessarily. All missing DBH values were given a value based on the average DBH of the same species for the year it was planted. Of the 28 species, only 3 species (in total 25 trees) were missing DBH values: 23 lemon (citrus), one balsa (timber), one bariable (other) trees. These were given DBH values interpolated from the other trees in the same family group and which were planted the same year.

Additionally, 8 banana trees that had DBH values larger than 50cm, which is unrealistically high. Assuming that there was a manual entry mistake, these values were exchanged with the maximum value of the banana trees for the year planted.

Aligning Drone Images with Field Boundaries

A key issue identified in the ReforesTree pipeline was the mismatch between the drone imagery boundaries and the field data boundaries. To address this, we implemented the following steps to align the drone images with the field measurements for the six agroforestry sites. The code for this is in this *reforestree-correction* repository.

1. **GeoDataframe Creation:** We converted the field data, which included the longitude and latitude of each point,
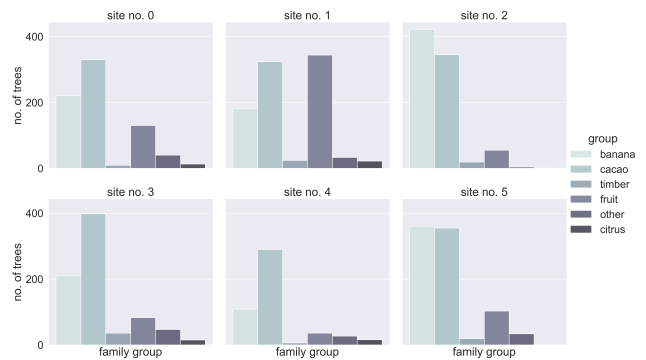

Figure 5: This figure represents the count of species family groups for each of the sites. All sites have trees of all species family groups, but cacao and banana are over represented.

into a GeoDataFrame using the geopandas library. This allowed us to create point geometries that were easy to visualize and manipulate. The field data points, visualized as red dots in Figure 6, served as the starting reference.

2. **Boundary Extraction using Alpha Shape:** To capture the boundary of the field data, we used the alphashape library to create a convex hull around the points. By choosing an alpha value of 15000, similar to the value used by (Barenne et al. 2022), we generated a tight boundary around the field data points.

3. **Overlay and Crop Drone Imagery:** Using the rasterio library, we overlapped the generated alphashape boundary onto the drone imagery (in TIFF format). We then cropped the unnecessary parts of the image, outside the boundary, replacing them with white pixels. This step is illustrated by the transition from the third to fourth images in Figure 6.

4. **Adjusting Image Boundaries:** Finally, after cropping, we identified the bounds of the non-white pixels in the images and adjusted them to ensure they fit a square shape correctly. This was essential for integrating the images into the AGBench library. The final result can be seen in the transition from the fourth to the last image in Figure 6.
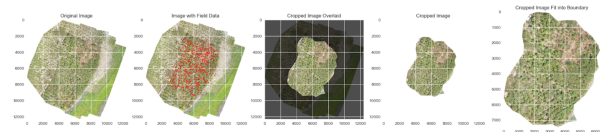

Figure 6: This figure shows the alignment process between the drone images and field boundaries. The field data points (red dots) were used to create the alphashape, which was overlaid onto the drone imagery to crop unnecessary areas and ensure accurate alignment.

Benchmark of satellite-based AGB maps

To benchmark the low resolution (LR) satellite-based maps, we fitted it to the high resolution (HR) drone imagery overlapping the GPS coordinates.
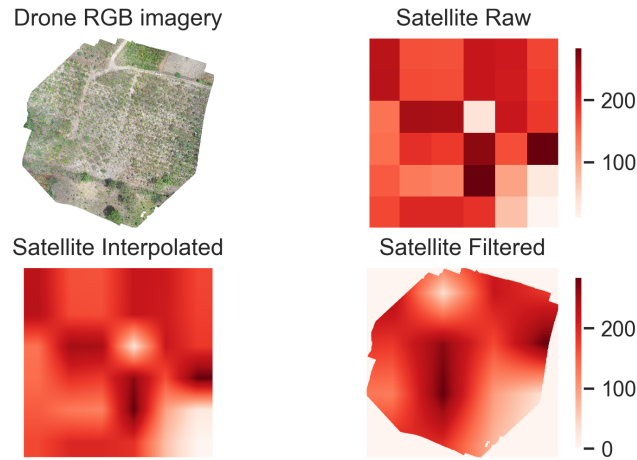

Figure 7: This figure represents the different steps in the benchmark analysis and how we calculated the total AGB amount from the satellite-based maps for the **ReforesTree** sites. This is taken from site no. 0. The values represented in the image is AGB density.

The calculation of the total AGB was done in five steps, illustrated in Figure 7

1. cropping the LR satellite map with a padding around the polygon of the site to reduce computation intensity (Satellite Raw)

2. linearly interpolating the values for this map and resize the map with the same HR pixel resolution as the drone imagery (Satellite Interpolated)

3. cropping the map further fitting with the GPS locations (max/min) of the drone imagery

4. filtering out the site area by removing all pixels in the satellite-based map, that are outside of the drone imagery, coloured white (Satellite Filtered)

5. lastly, multiplying the AGB mean density of the filtered map with the project site area to get the total AGB

We analysed the following three maps:

- (Global Forest Watch 2019): Aboveground Woodly Biomass with 30x30m resolution for the year of 2000.

- (Spawn, Sullivan, and Lark 2020): Global Aboveground and Belowground Biomass Carbon Density Maps for the Year 2010 with 300x300m resolution.

- (Santoro 2018): GlobBiomass - Global Datasets of Forest Biomass with 100x100m resolution for the year 2010.

Baseline CNN

We trained the model on a single GPU of the type GeForce RTX 3090. The learning rate used was $1e^{-3}$, batch size of 64 for 30 epochs achieving a root square mean loss (RMSE) of 0,1.