Forests & Land Systems

CCAI Summer School 2024

David Dao, PhD

Co-Founder @ Gainforest.Earth

Co-Lead @ ETH BiodivX

Gainforest.Earth

About me


BIENVENIDOS
PARQUE NACIONAL
DEFENSORES DEL CH
MADES


COP25
HILE
ID 2019
HANGE CONFERENCE

28 Partner Organisations in the Global South

Kenya

Brazil

Philippines

Lecture Overview

1. Climate Change: Forests and Land Systems
2. Saving Forests: Finance and Regulations
3. Artificial Intelligence: How can it help?
4. Equitable AI: The Role of IPLCs and Frontier Data
5. Future Work & Discussion

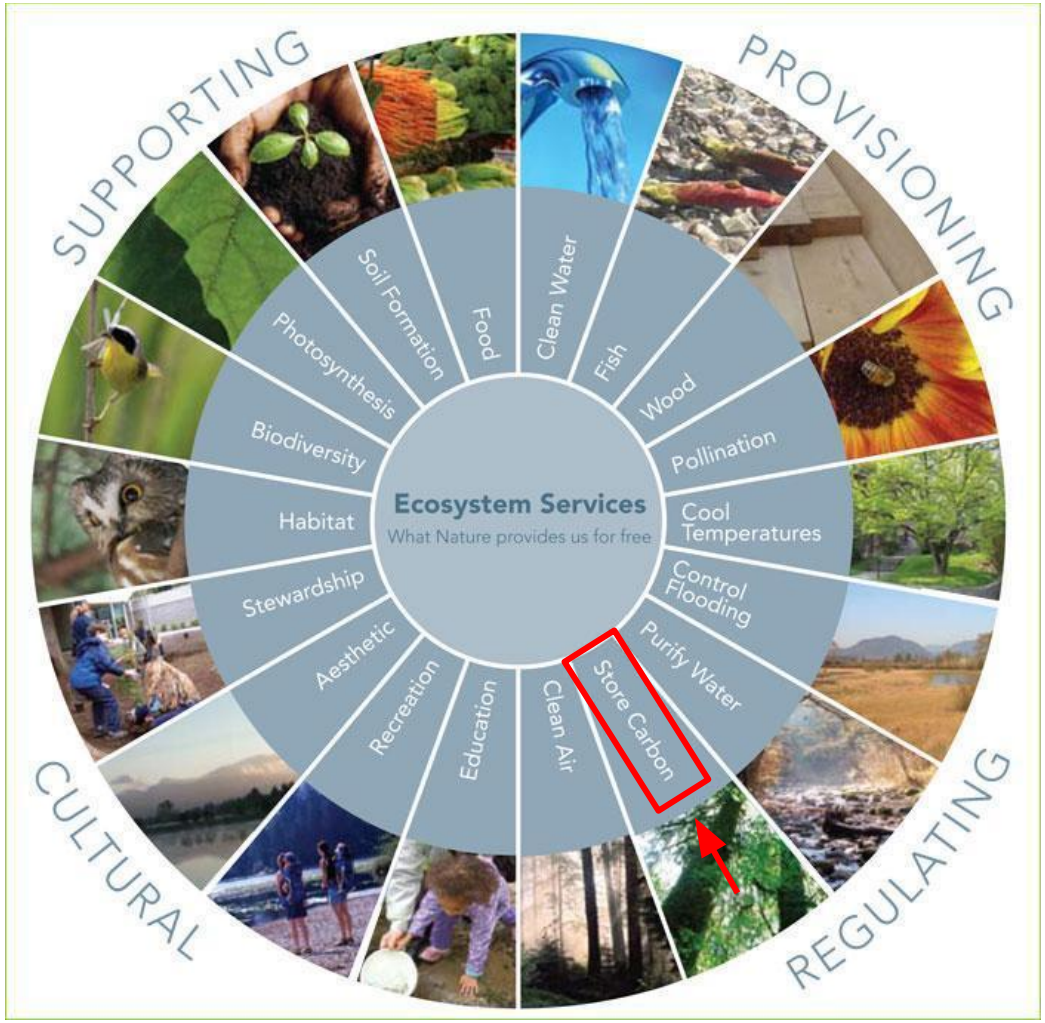
Lecture Overview

1. Climate Change: Forests and Land Systems
2. Saving Forests: Finance and Regulations
3. Artificial Intelligence: How can it help?
4. Equitable AI: The Role of IPLCs and Frontier Data
5. Future Work & Discussion

Listen 🌱

“Nature is essential for human existence and good quality of life. Most of nature’s contributions to people are not fully replaceable, and some are irreplaceable.”

The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services

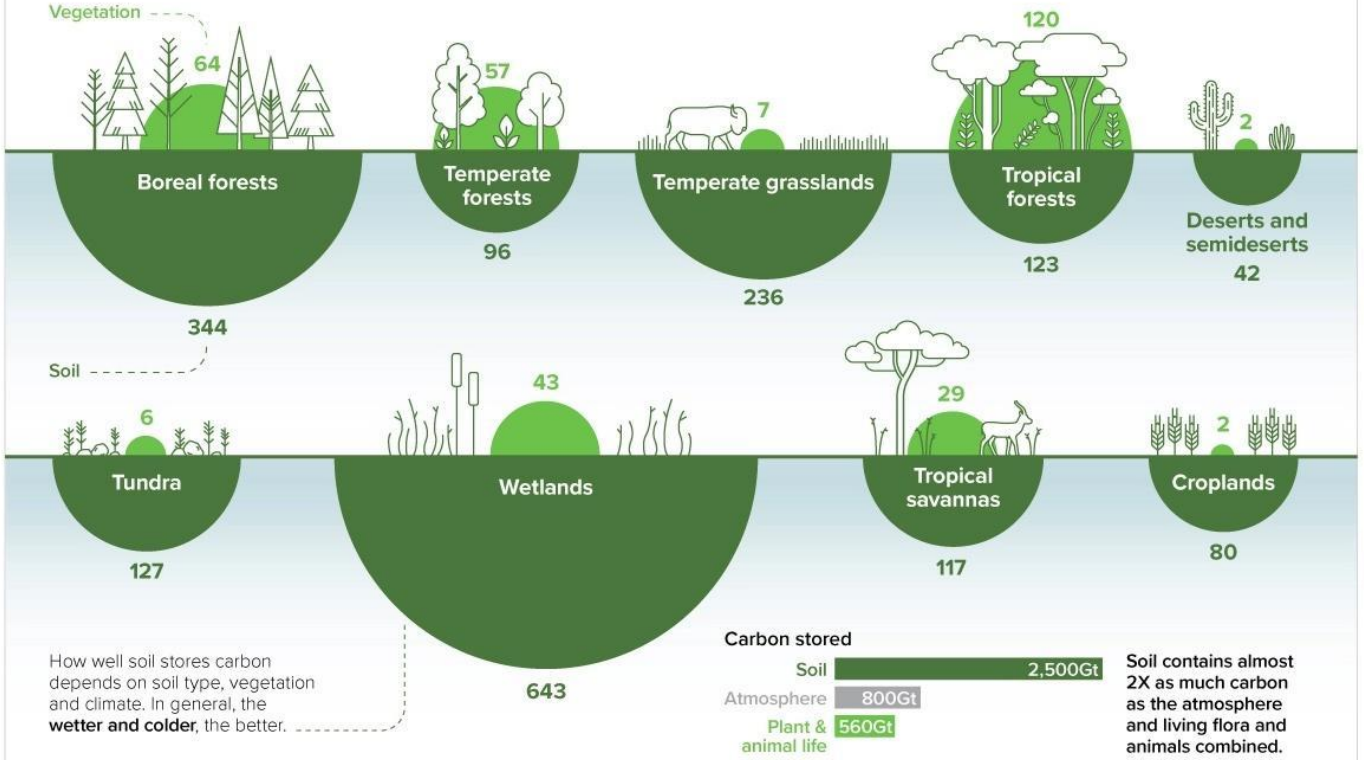

SUPPORTING
PROVISIONING
CULTURAL
REGULATING
Ecosystem Services
What Nature provides us for free
Food
Soil Formation
Photosynthesis
Biodiversity
Habitat
Stewardship
Aesthetic
Recreation
Education
Clean Air
Store Carbon
Purify Water
Control Flooding
Cool Temperatures
Pollination
Wood
Fish
Clean Water

Carbon Storage

Tonnes of Carbon per Hectare*

The world's forests absorb around **15.6 gigatonnes** of CO_2 each year. That's around 3X the annual CO_2 emissions of the United States.

However, around **8.1 gigatonnes of CO_2 leaks back into the atmosphere** due to deforestation, fires and other disturbances.



How well soil stores carbon depends on soil type, vegetation and climate. In general, the **wetter and colder**, the better.

Carbon stored

Soil contains almost 2X as much carbon as the atmosphere and living flora and animals combined.

*At a ground depth of one meter

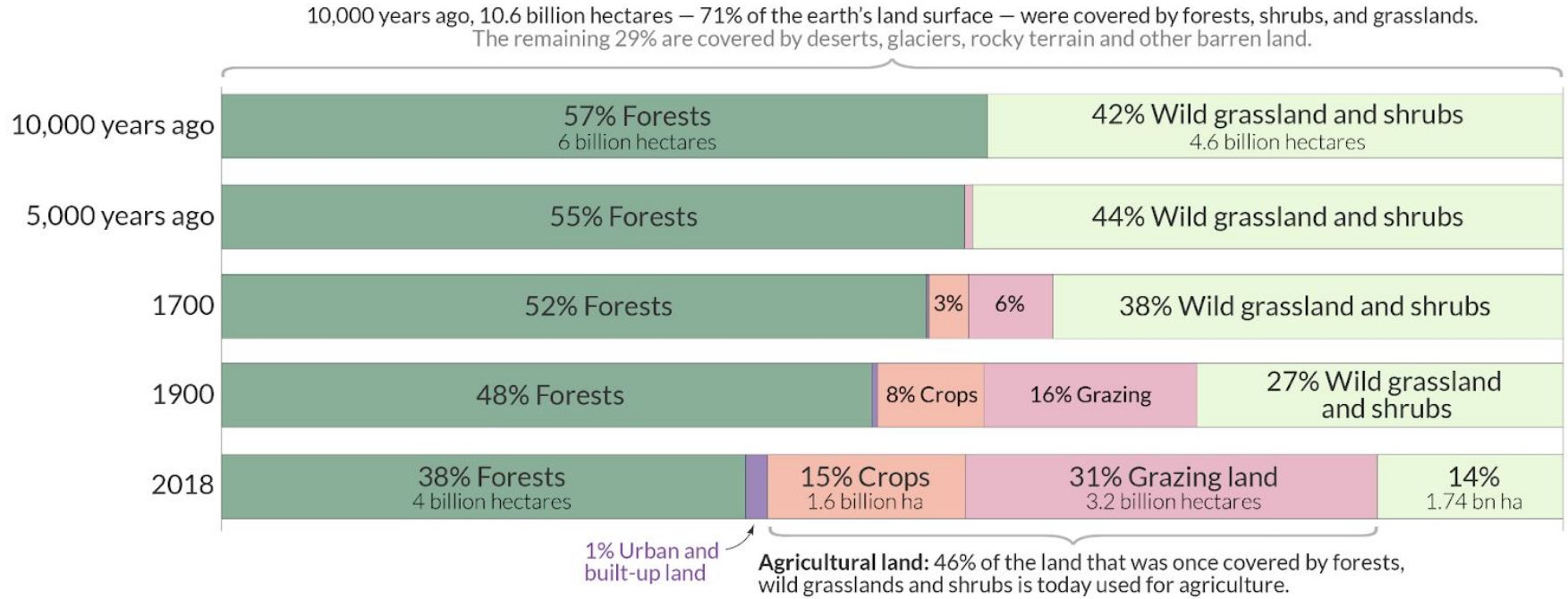
Sources: IPCC; NASA


Location
Mato Grosso, Brazil
1984
“The natural world is deteriorating in rates unparalleled in human history”
The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services
Google Earth

Humanity destroyed one third of the world's forests by expanding agricultural land

Agriculture is by far the largest driver of deforestation. To bring deforestation to an end humanity has to find ways to produce more food on less land.

Our World in Data

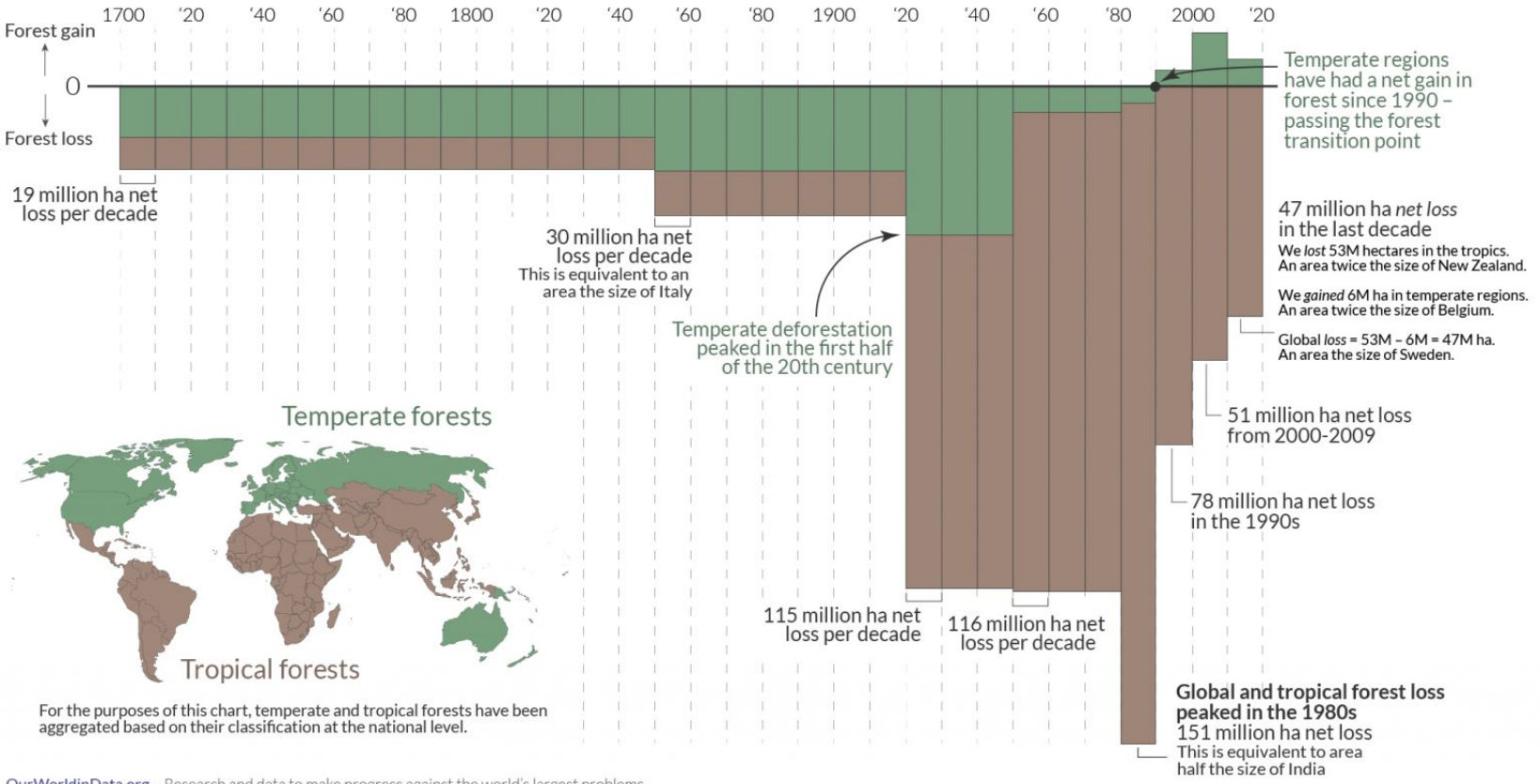


Data: Historical data on forests from Williams (2003) – Deforesting the Earth. Historical data on agriculture from The History Database of Global Environment (HYDE). Modern data from the FAO.
OurWorldinData.org – Research and data to make progress against the world's largest problems.

Decadal losses in global forest over the last three centuries

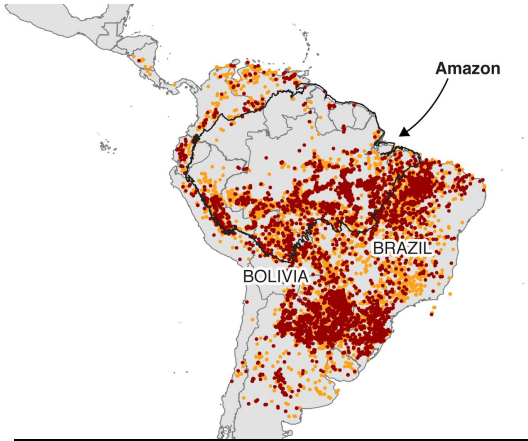
Decadal forest loss is measured as the average net loss of forest area every ten years, in hectares.
This equals deforestation minus any increases in forest area through afforestation.
1.5 billion hectares of global forest was lost between 1700 and 2020 – this is equal to an area 1.5-times the size of the USA.

Our World in Data



OurWorldinData.org – Research and data to make progress against the world's largest problems.
Data sources: Based on pre-1990 data from Williams (2006). Deforesting the Earth. Post-1990 data from UN FAO Global Forest Resources Assessment (2020).

Deforestation is a key driver of the climate crisis



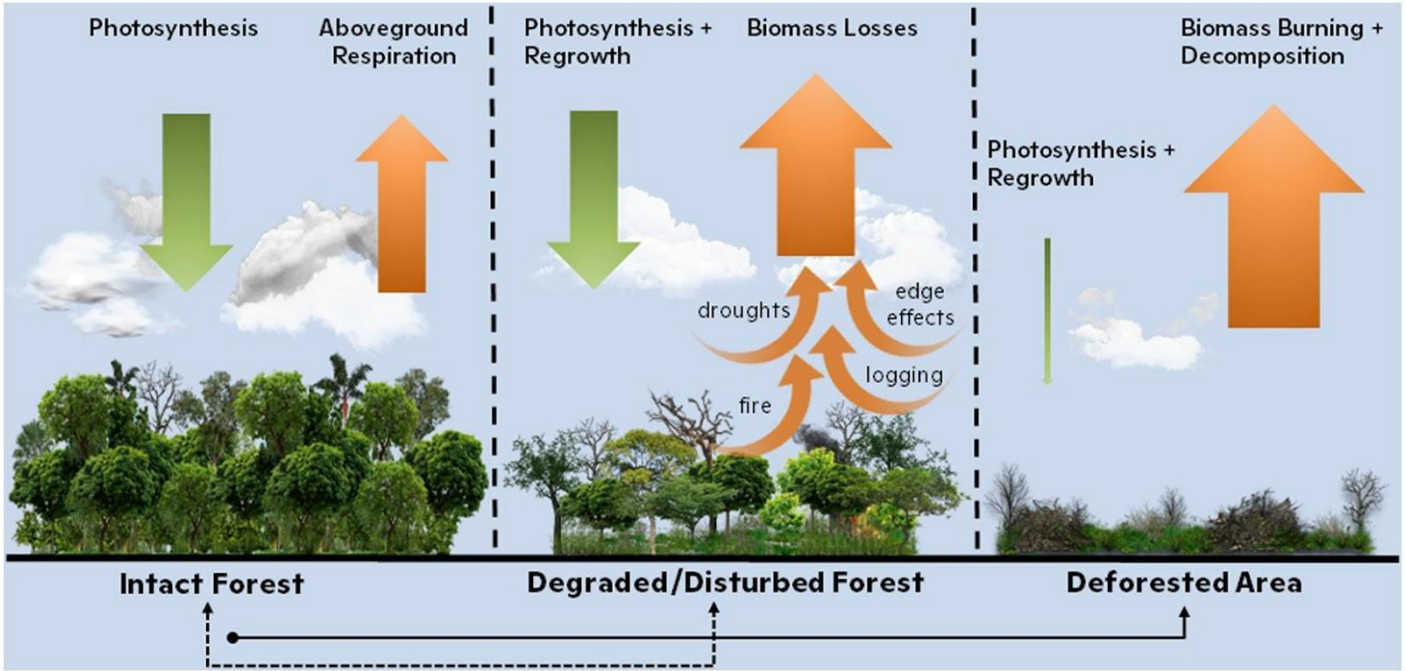
Through deforestation, the Amazon rainforest turned from a carbon sink into a carbon source

Source: MODIS, Planet Labs imagery

18%

Global anthropogenic emissions

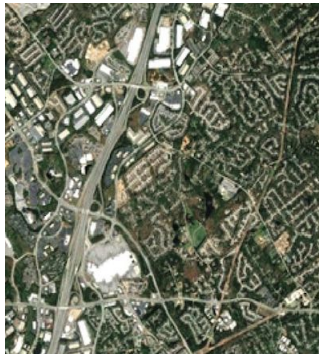
Deforestation and forest degradation



Beyond Deforestation: Carbon Emissions From Land Grabbing and Forest Degradation in the Brazilian Amazon, *Kruid et al.* Frontiers https://doi.org/10.3389/ffgc.2021.645282

Drivers of Forest Loss (Quiz ❓)

a)



b)

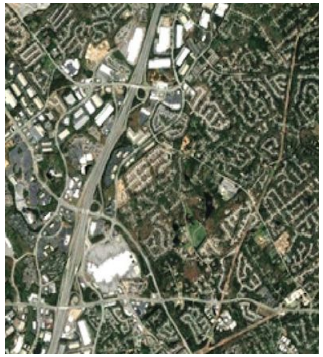
c)

d)

e)

Drivers of Forest Loss (Quiz ❓)

Deforestation
(27%)



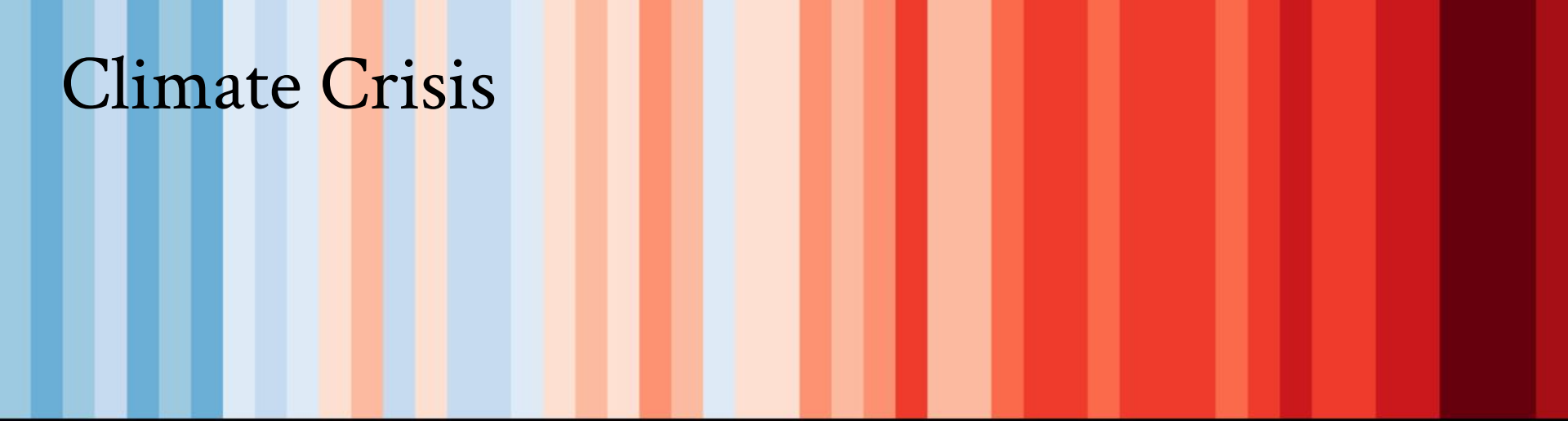
Urbanization
(1%)

Shifting agriculture
(24%)

Forest Products
(26%)

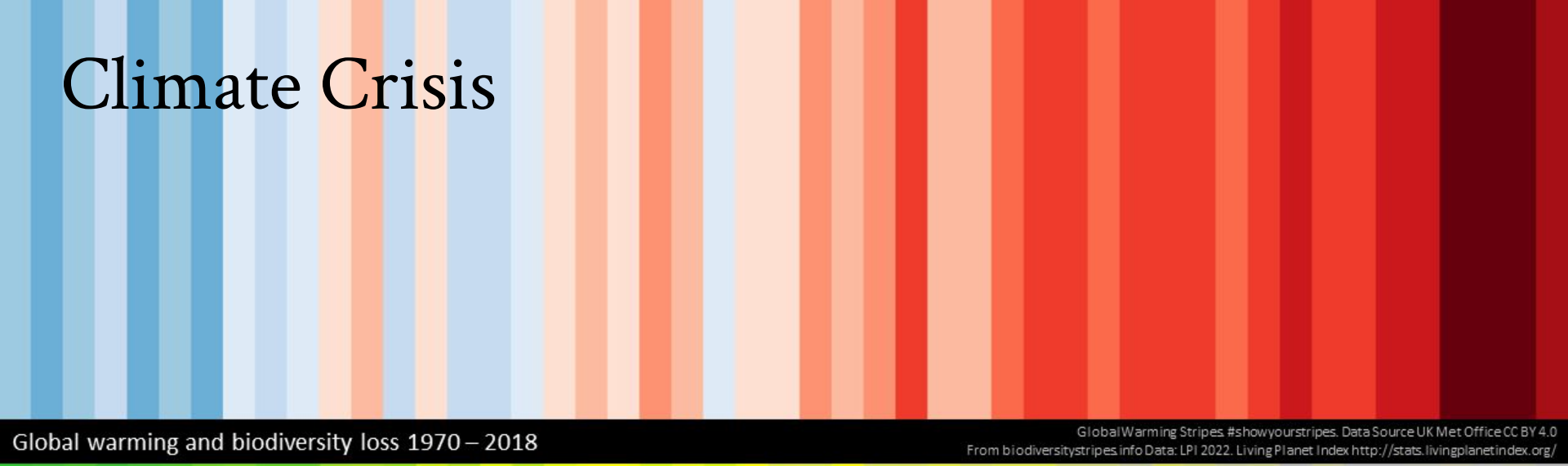
Wildfires
(23%)

Climate Crisis



Global warming 1970 – 2018

Climate Crisis



Biodiversity Loss

“Deforestation is changing our climate, harming people and the natural world. We must, and can, reverse this trend.”

Dr. Jane Goodall

One Trillion Trees Reforestation Potential

10%

Of emission cuts needed to stay below 1.5 deg*

Global Reforestation Potential, Science 2019, Bastin *et al.*

*rough estimate (but *we* all agree that restoration done right has incredible benefits)

Lecture Overview

1. Climate Change: Forests and Land Systems
2. Saving Forests: Finance and Regulations
3. Artificial Intelligence: How can it help?
4. Equitable AI: The Role of IPLCs and Frontier Data
5. Future Work & Discussion

Forest Carbon Credits

Reforestation

Planting young trees to remove carbon

Estimated Impact: 15 tonnes of CO2 per hectare per year

Improved Forest Management (IFM)

Preserve middle-aged trees (captures and stores carbon)

Avoided Deforestation

Preserve old-growth forest

Estimated Impact: 400 tonnes of CO2 per hectare *(average Amazon rainforest)*

Nature Finance

Payment for Ecosystem Services (PES)

Payment for Ecosystem Services (PES) is a concept where landowners are compensated for managing their land in a way that provides some sort of ecological service.

Challenge: Land rights, finding the right KPIs, monitoring

Agroforestry

Agroforestry is a land use management system in which trees or shrubs are grown around or among crops or pastureland.

Challenge: Market access and extensive domain expertise

Ecotourism

Ecotourism is a form of tourism that is focused on the conservation of nature and the well-being of local people.

Challenge: No overuse and equitable share of benefits to communities

Nature Regulations

EU Deforestation Law

The EU Deforestation Law aims to curb the EU's contribution to global deforestation and forest degradation. It mandates that companies ensure their products do not contribute to deforestation, covering commodities like soy, palm oil, wood, cocoa, coffee, and beef.

Challenge: Verifying the deforestation-free status of products

Global Biodiversity Goals

The agreement sets ambitious targets, such as protecting 30% of the planet's land and sea areas by 2030.

Challenge: Establishing effective mechanisms for monitoring progress and ensuring accountability

EU Nature Restoration Act

The EU Nature Restoration Act aims to restore degraded ecosystems across the EU, with legally binding targets to restore at least 20% of the EU's land and sea areas by 2030 and all ecosystems in need of restoration by 2050.

Challenge: Ensuring compliance with restoration targets and preventing further degradation

Monitoring, Reporting and Verification (MRV)

or “We cannot value what we cannot measure”

Results-based payments require monitoring, reporting and verification (MRV)

$25bn

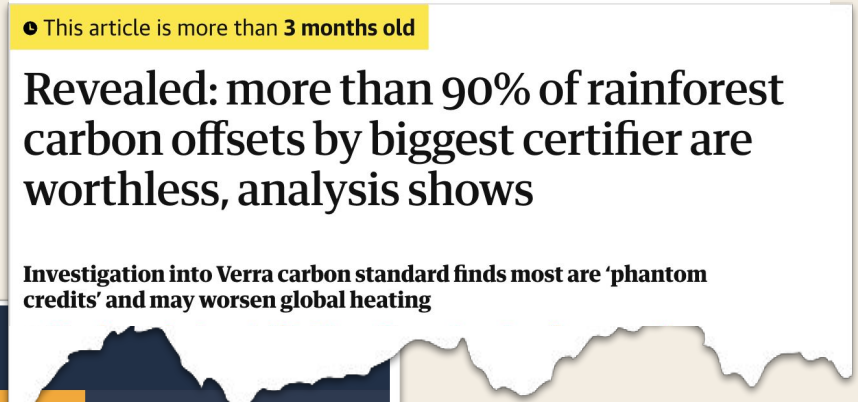
Public funds reserved for results-based payments in forestry

But overlapping land claims, slow performance measurement and missing trust

Make funds inaccessible to local communities

Source: REDD+

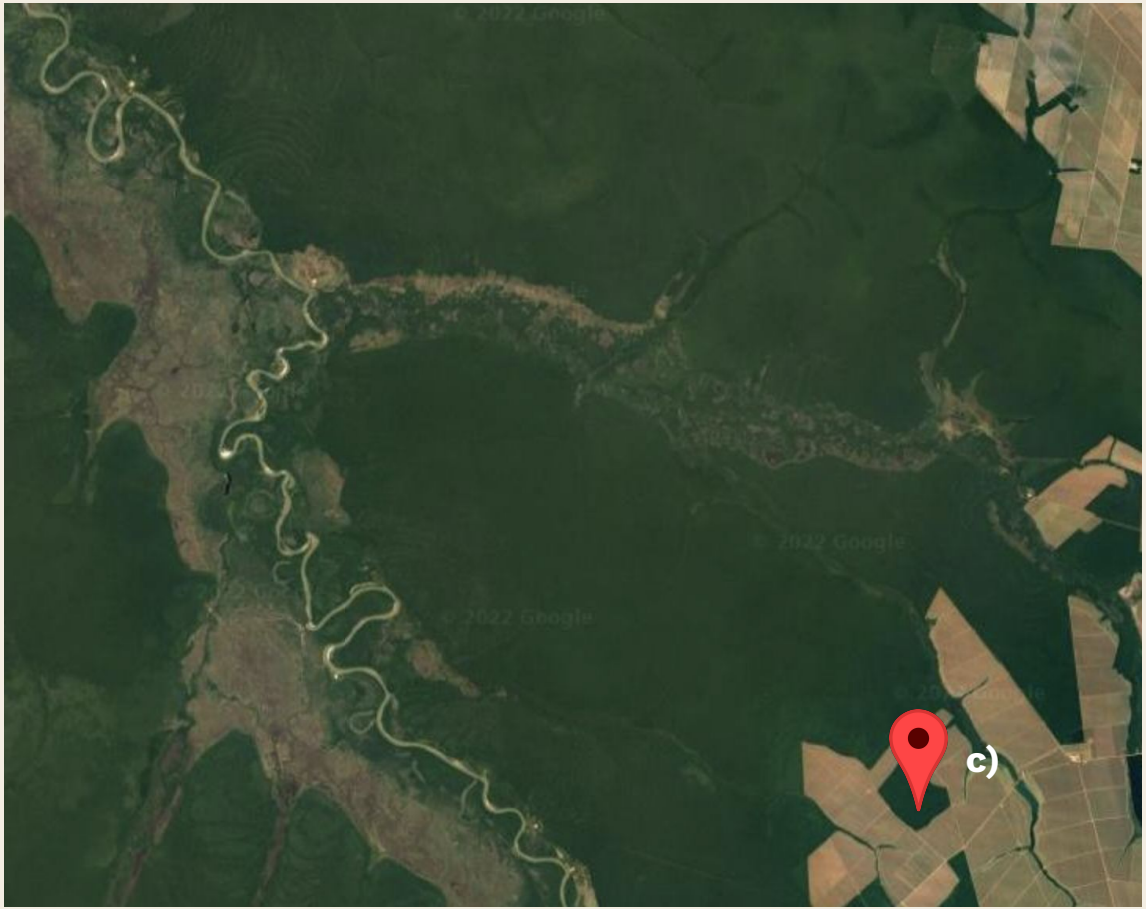
Is it greenwashing?



Additionality? (Quiz ❓)

Additionality: What was the risk of deforestation in the first place?

Leakage?



Leakage: Maybe c) was protected but owner just cashed in and instead deforested surrounding

Listen again 😰


Co-Benefits?

Important Note

Carbon Credits

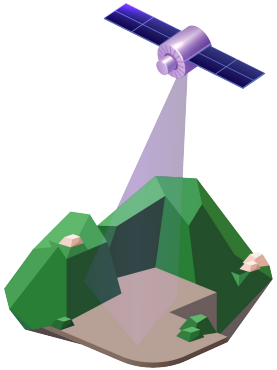
without safeguards can harm climate and environment

Lecture Overview

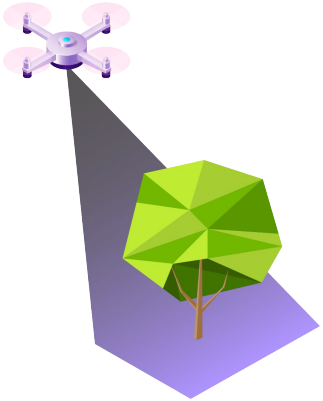
1. Climate Change: Forests and Land Systems
2. Saving Forests: Finance and Regulations
3. Artificial Intelligence: How can it help?
4. Equitable AI: The Role of IPLCs and Frontier Data
5. Future Work & Discussion

Explosion of multi-modal environmental data

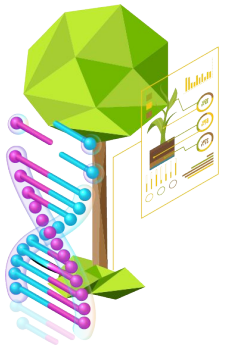
Field-based monitoring



Satellite-based monitoring



Drone-based monitoring



eDNA monitoring

How can we leverage this data?

Field-based monitoring

Satellite-based monitoring

Drone-ba monitor

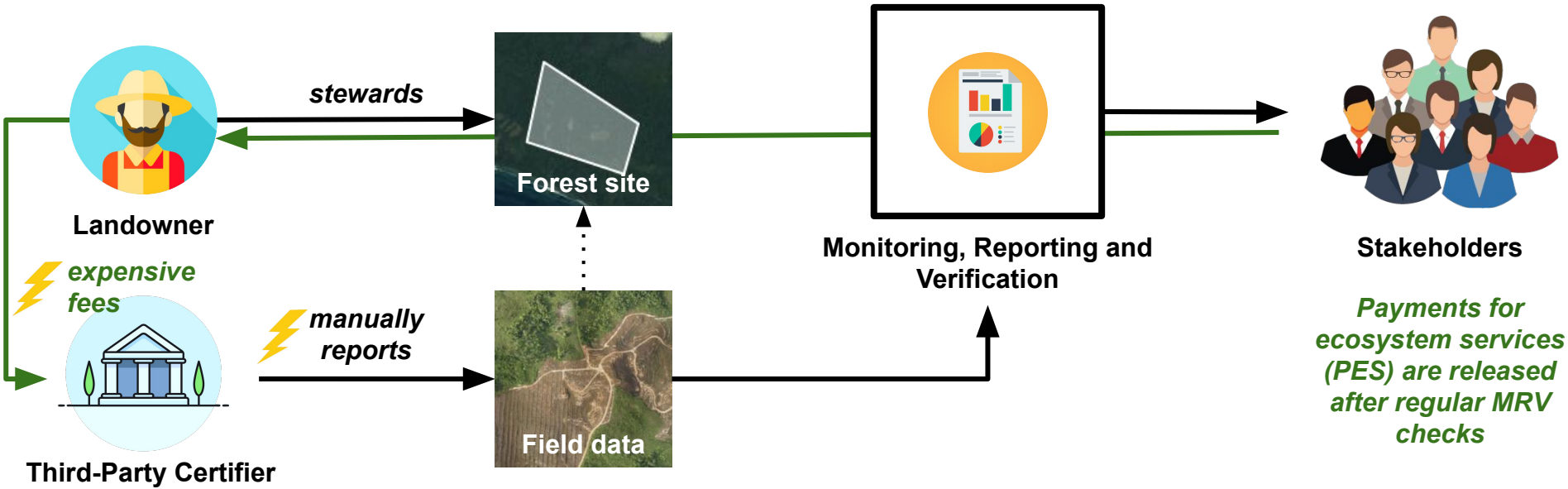
eDNA monitoring

Fusing and learning from multi-modal environmental data provide us with novel *opportunities* (and *challenges)*

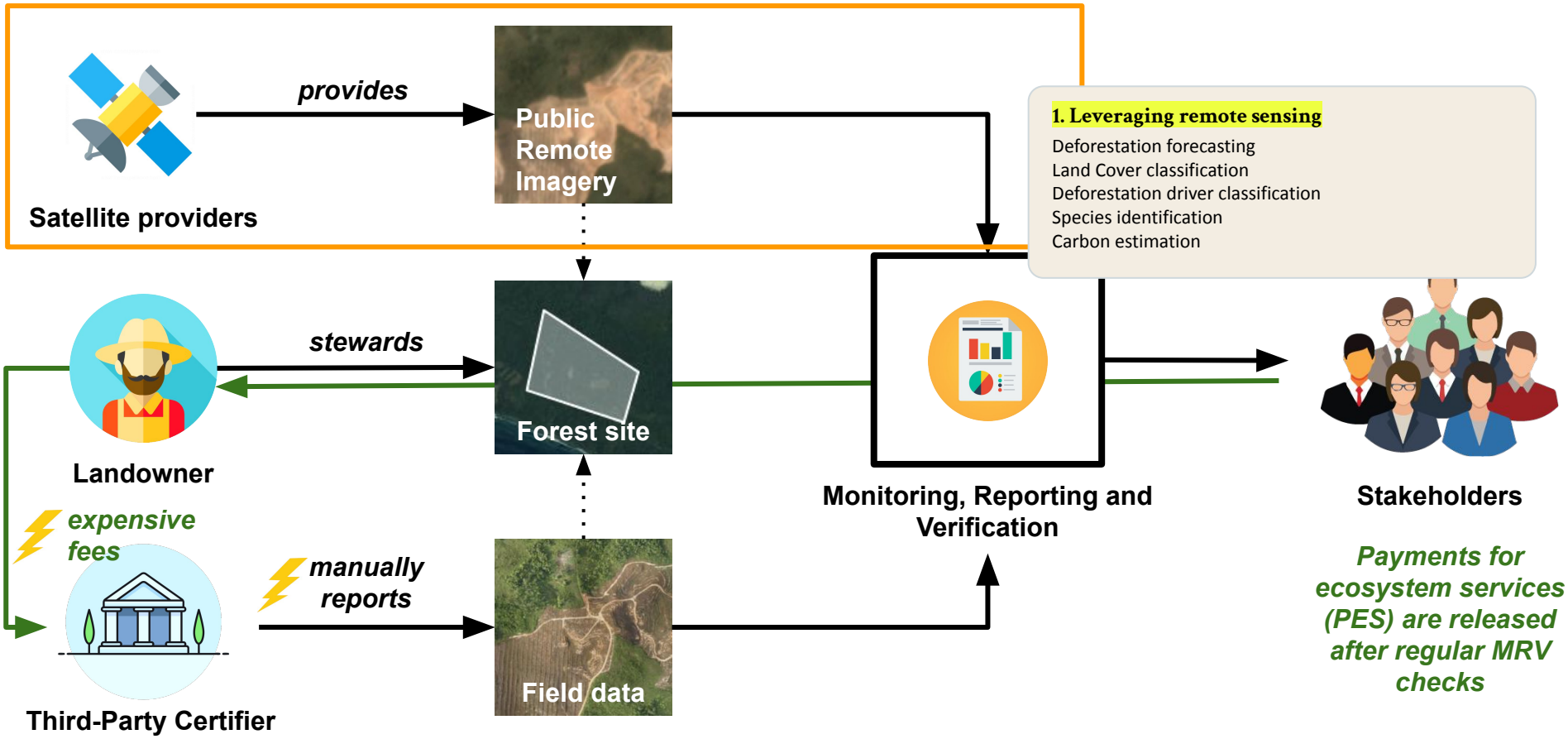
AI is uniquely positioned to do so!

“ML can help us understand the location, health and ecological value of nature and biodiversity at scale, and ensure these metrics are reflected in policy, finance, and decision-making.”

Overview


stewards
Forest site
Landowner
expensive fees
manually reports
Field data
Third-Party Certifier
Monitoring, Reporting and Verification
Stakeholders
Payments for ecosystem services (PES) are released after regular MRV checks

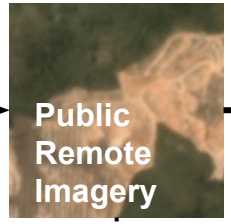
Overview


provides
Public Remote Imagery
Satellite providers
1. Leveraging remote sensing
Deforestation forecasting
Land Cover classification
Deforestation driver classification
Species identification
Carbon estimation
stewards
Forest site
Landowner
Monitoring, Reporting and Verification
Stakeholders
expensive fees
manually reports
Field data
Third-Party Certifier
Payments for ecosystem services (PES) are released after regular MRV checks

Overview

Satellite providers

provides



1. Leveraging remote sensing

Deforestation forecasting
Land Cover classification
Deforestation driver classification
Species identification
Carbon estimation

Landowner

stewards

Monitoring, Reporting and Verification

Stakeholders

Payments for ecosystem services (PES) are released after regular MRV checks

low fees

Third-Party Certifier

machine-assisted reporting

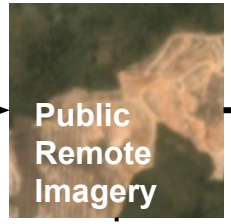
2. Connecting with field data

A mobile data app to connect field with satellite data
Measure biodiversity using ecoacoustic

Overview

Satellite providers

provides



1. Leveraging remote sensing

Deforestation forecasting
Land Cover classification
Deforestation driver classification
Species identification
Carbon estimation

Landowner

stewards

3. Decentralize MRV

Empower everyone to submit forest data
For verification

rewards

Citizen science

machine-assisted reporting

2. Connecting with field data

A mobile data app to connect field with satellite data
Measure biodiversity using ecoacoustic

Monitoring, Reporting and Verification

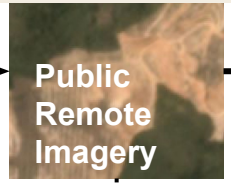
Stakeholders

Payments for ecosystem services (PES) are released after regular MRV checks

Can we leverage remote sensing for MRV?

Satellite providers

provides



Landowner

stewards

Monitoring, Reporting and Verification

Stakeholders

Payments for ecosystem services (PES) are released after regular MRV checks

Utilize multi-modal imagery sources

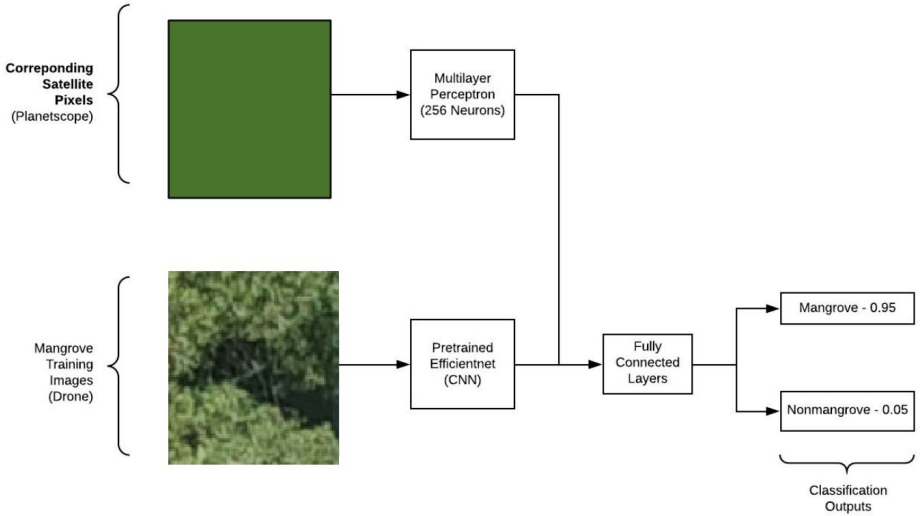
Left: Public data
(Sentinel-2, LANDSAT-8)
>10m/px

Right: Commercial data
(Planet Labs, MAXAR)
>50cm/px

Challenges:
Different sensors, reference systems, access levels and temporal info

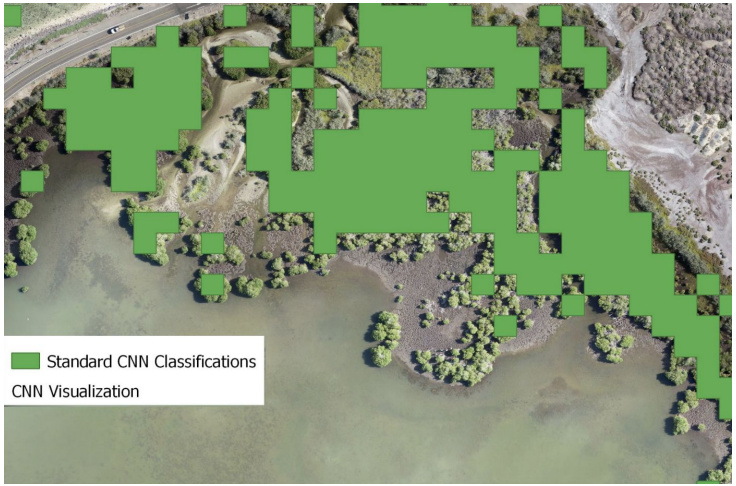
Estância Shalom, Brazil

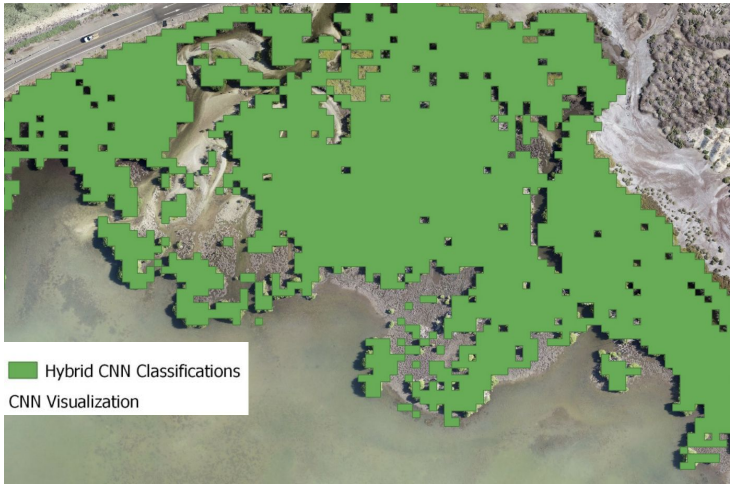
Mangrove classification



Fusing different data resolutions improves classification accuracy

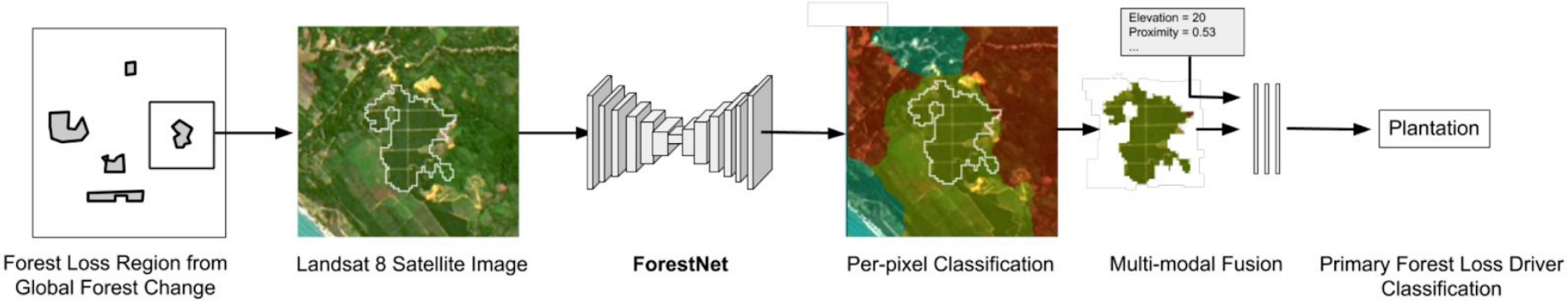
Challenges: Labeled Data





Mangrove Ecosystem Detection using Mixed-Resolution Imagery with a Hybrid-Convolutional Neural Network
Hicks *et al.*, CCAI ICML20

Deforestation driver classification

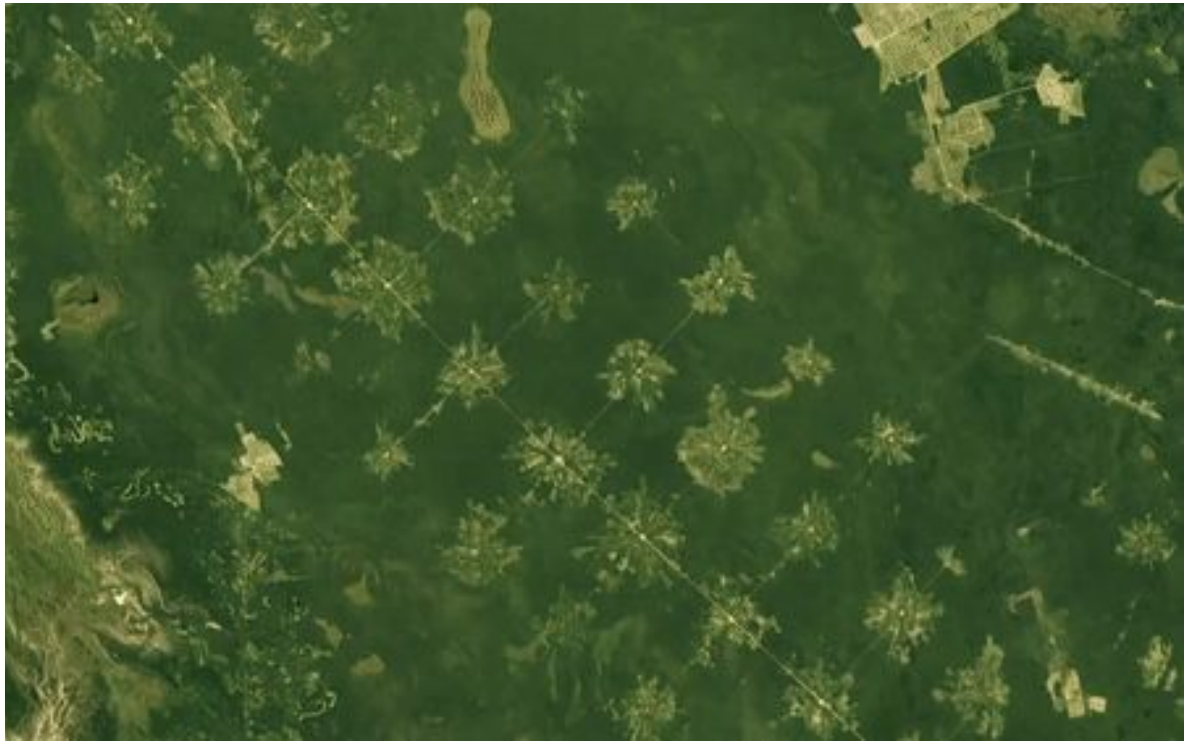


Model	Predictors	Val		Test	
		Acc	F1	Acc	F1
RF	Visible	0.56	0.49	0.49	0.44
RF	Visible + Aux	0.72	0.67	0.67	0.62
CNN	Visible	0.80	0.75	0.78	0.70
CNN + SDA	Visible	0.82	0.79	0.78	0.73
CNN + SDA + PT	Visible	0.83	0.80	0.80	0.74
CNN + SDA + PT	Visible + Aux	**0.84**	**0.81**	**0.80**	**0.75**

Fusing multi-modal data sources improve classification accuracy
Challenges: Labeled Data

ForestNet: Classifying Drivers of Deforestation in Indonesia using Deep Learning on Satellite Imagery
Irvin *, Sheng * *et al.* CCAI NeurIPS20

Utilize large time series

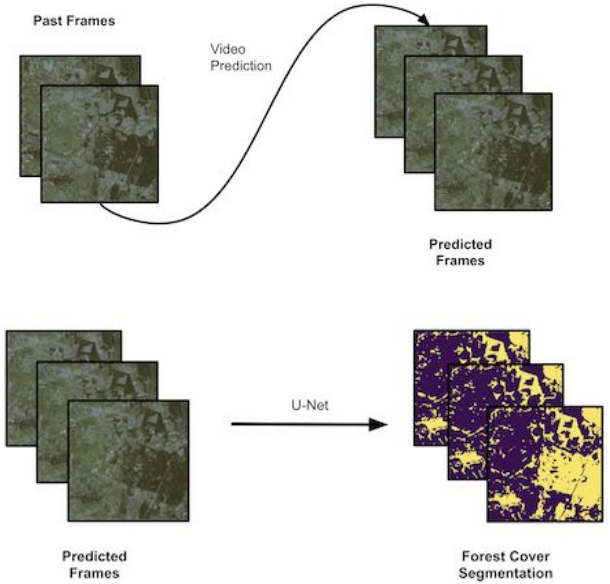


Challenges:

Largely unlabeled data:
E.g. deforestation exhibits visually recognizable patterns

Irregular temporal steps:
Visual data depends on satellite revisiting rate and cloudfree image

Deforestation forecasting with self-supervision



Challenges:
Difficult to train
Needs lots of computational resources

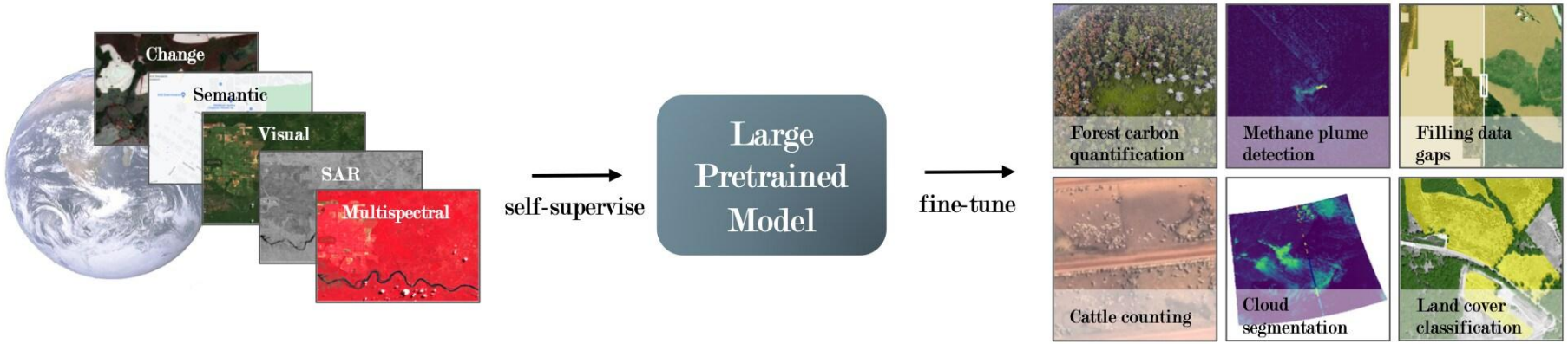
Landsat Time Series

Forest Loss (Hansen)

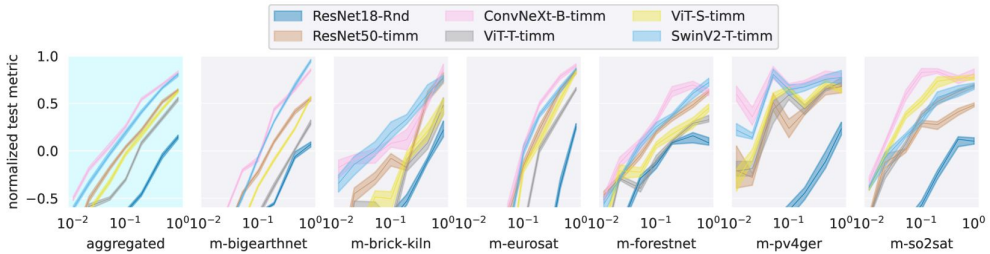
Video Prediction

Video Prediction + U-Net

Geospatial Foundation Models

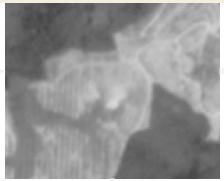


Challenges:
Difficult to train
Needs lots of computational resources



GEO-Bench: Toward Foundation Models for Earth Monitoring
Lacoste et *al.*, NeurIPS Benchmark 23

What is the role of field-based ML?



Landowner

stewards

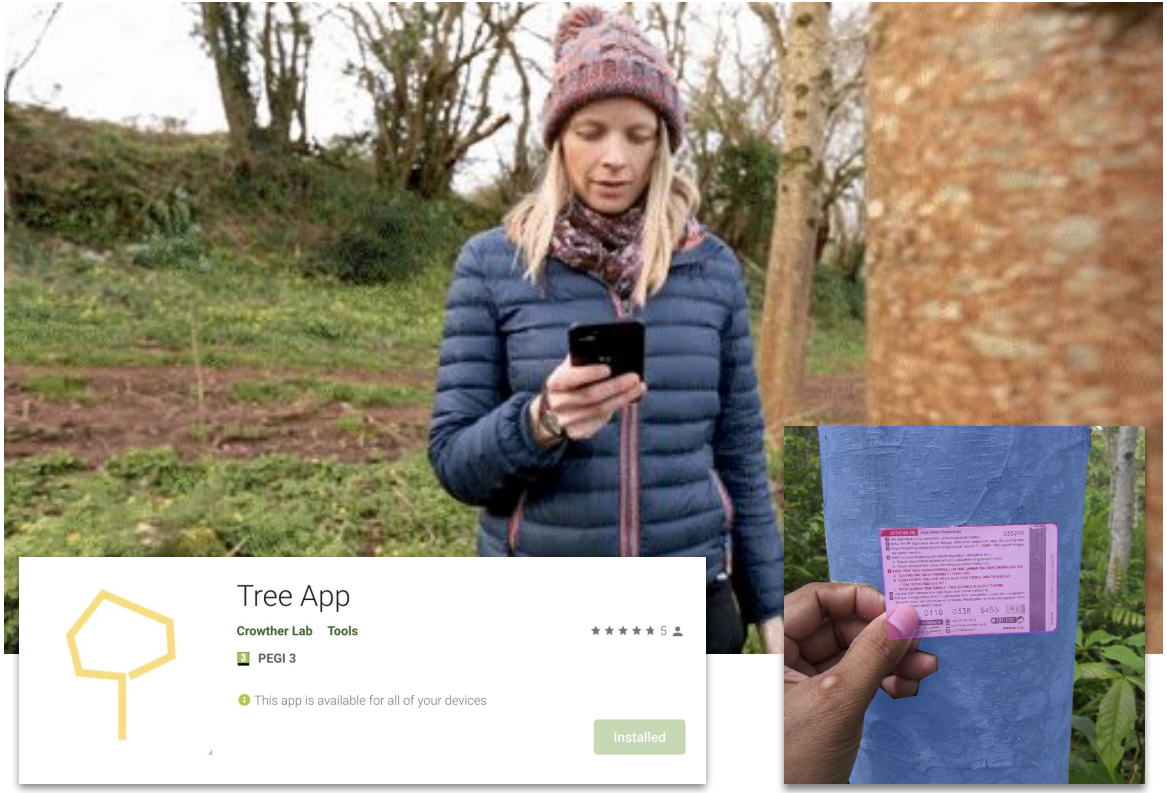
Monitoring, Reporting and Verification

Stakeholders

machine-assisted reporting

Payments for ecosystem services (PES) are released after regular MRV checks

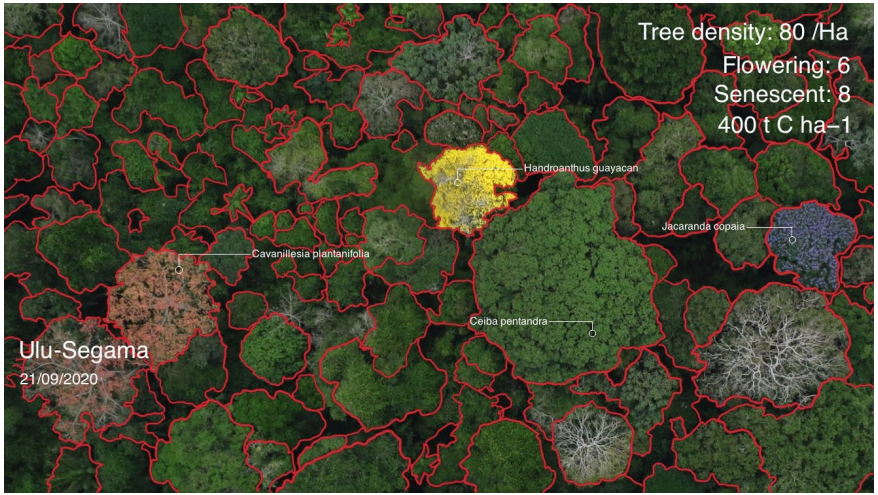
Machine-Learning Based Allometric Estimation

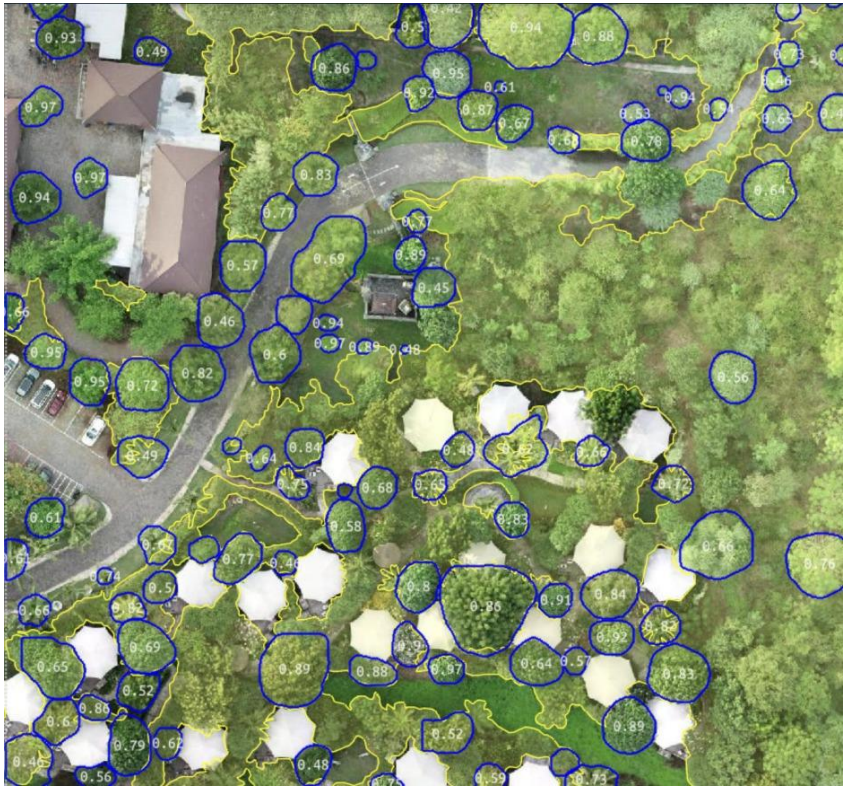


Tree App:
Estimating diameter at breast height (DBH) and species automatically from a single image

Challenges:
Noisy GPS
Low-cost and offline models
Collecting data still tedious

Drone-Based Biomass Estimation





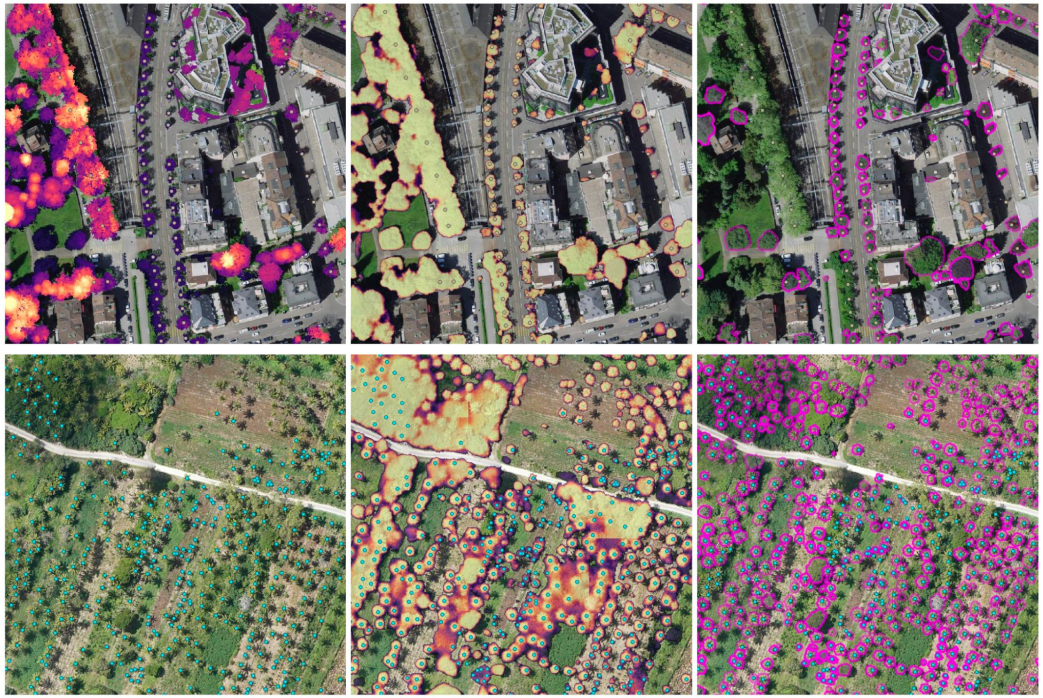
Challenges:
Limited labeled data for model training

Tree Canopy Delineation

Challenges:
Instance Segmentation vs Canopy Segmentation

Zurich

Tonga



Ground Truth | Canopy Seg | Instance Seg

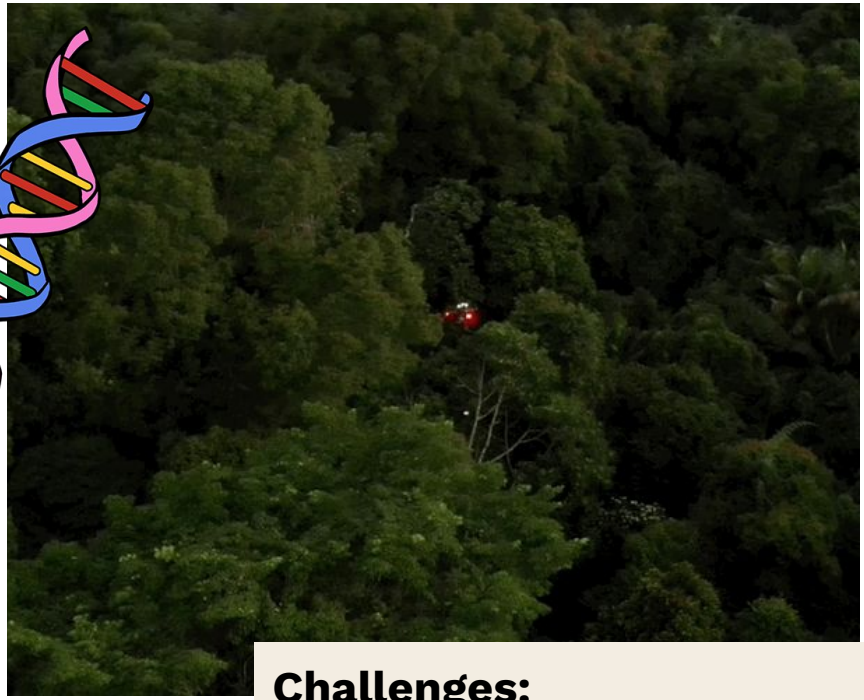
In Submission (Veitch *et al.*)

Drone-Based Biodiversity Measurements

Work from our collaborator Prof. Stefano Mintchev in “Environmental Robotics”

Challenges:
Difficult Terrain
Network Connection

Drone-Based Environmental DNA Sampling



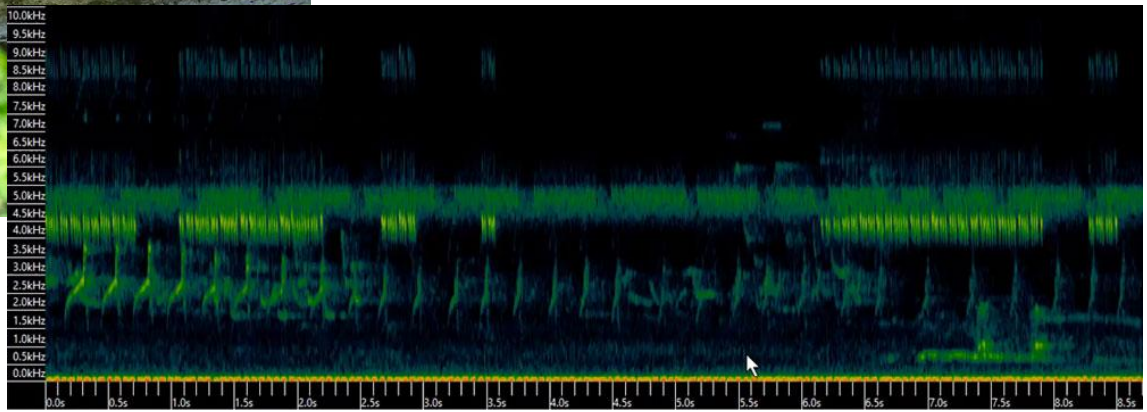
Sample Strategies:
eDNA through water filtration
eDNA through surface collection
eDNA through air filtration

Challenges:
Collecting enough eDNA
Dense Canopy

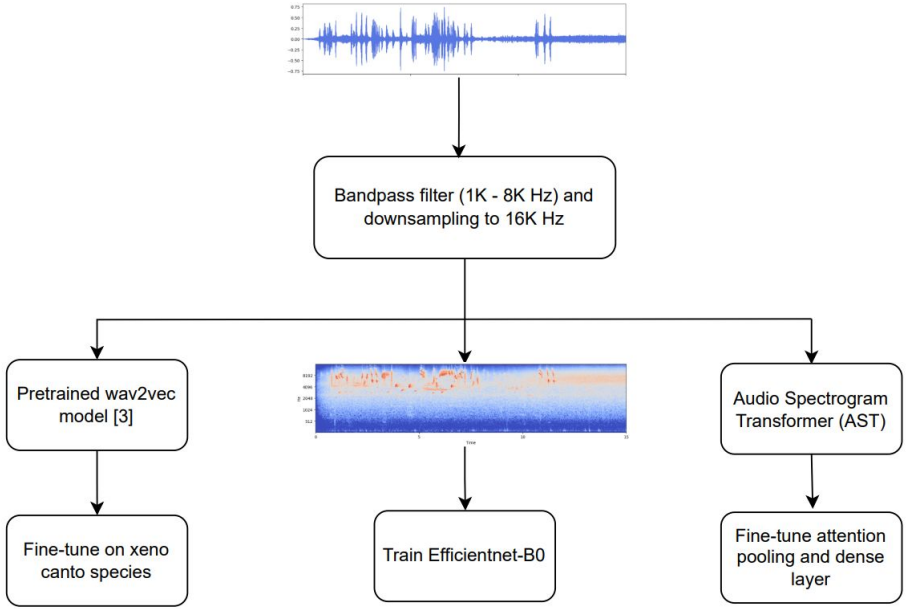
Joint work displayed during the XPRIZE semifinal in Singapore (Deiner, Mintchev, Dao *et al.*)

Bioacoustic Sensors

Audiomoths are low price and can record 24/7 for two weeks with 3 AA Batteries



Species Predictions



Easy to deploy and wide consistent range (500m)

Challenges: Labeled Data

Visualisation credits to **VOJ**

Soundscapes (Quiz ❓)

a)

b)

c)

d)

Soundscapes (Quiz ❓)

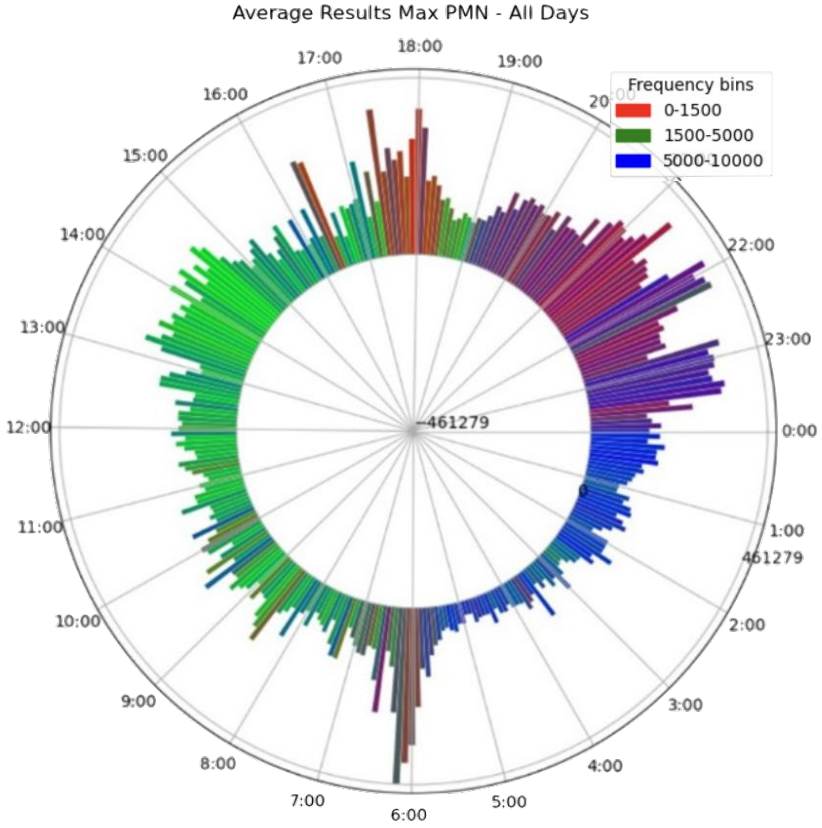
Reference Forest

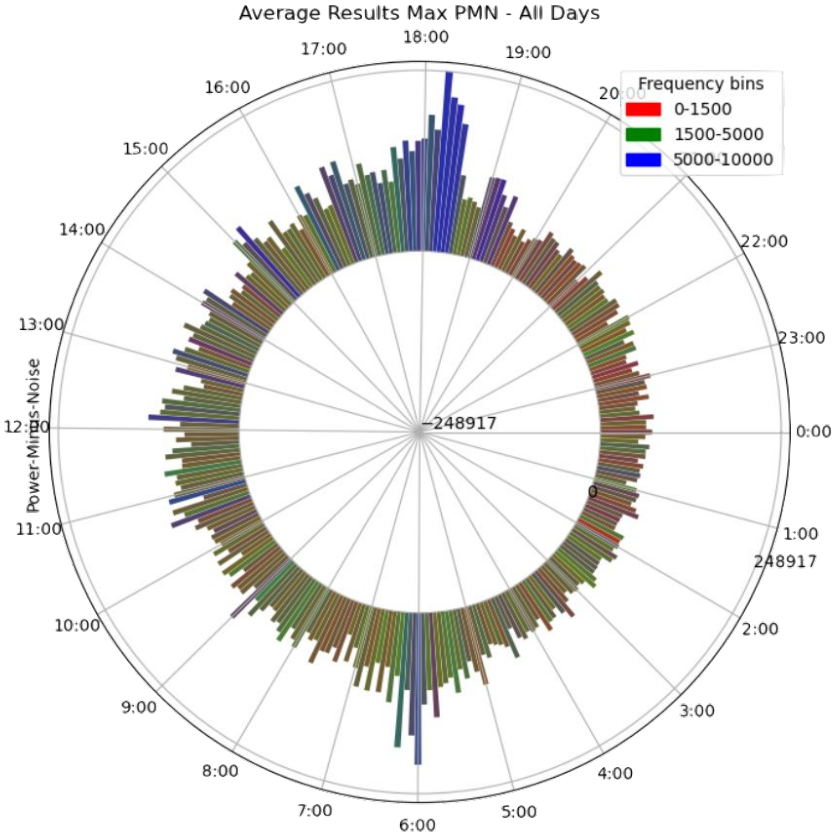
Natural Regeneration

Plantation

Pasture

Soundscapes





Visualisation credits to **confused scientists**

Lecture Overview

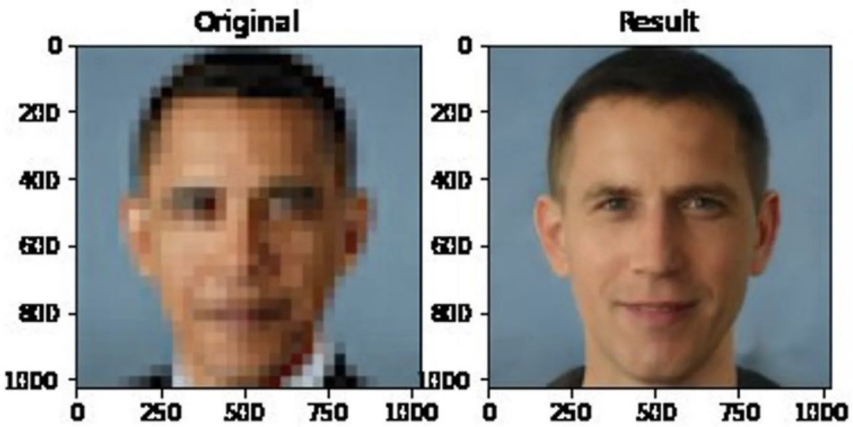
1. Climate Change: Forests and Land Systems
2. Saving Forests: Finance and Regulations
3. Artificial Intelligence: How can it help?
4. Equitable AI: The Role of IPLCs and Frontier Data
5. Future Work & Discussion

Limitations of AI: Bias

Our data is biased.
Algorithms can amplify existing systematic biases.

Source: The Algorithmic Justice League, MIT Media Lab

Limitations of AI: Bias



Source: https://github.com/daviddao/awful-ai

Countries resized for their available biodiversity data

Sampling biases shape our view of the natural world
Alice C. Hughes, Michael C. Orr, Keping Ma, Mark J. Costello, John Waller, Pieter Provoost, Qinmin Yang, Chaodong Zhu, Huijie Qiao

Where are the study sites?

Collaboration with local communities in Philippines and Ecuador

Lobo, Batangas
approx. 20 hectares

Loma Linda, Ecuador
6 sites, each approx. 2 hectares

Systematic Overestimation Study

- AGB dataset from agro-forestry sites
- 4663 trees, 28 species and 3.17 ha
- Each tree registered with DBH, species, and GPS location
  RGB drone images per site

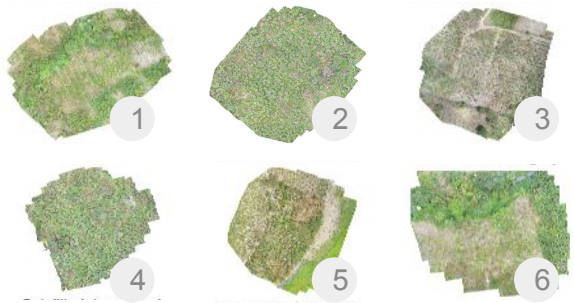


Fig. 1 Information about each site

SITE NO.	NO. OF TREES	NO. OF SPECIES	PLOT AREA	AGB DENSITY
1	743	17	0.53	19
2	929	19	0.47	32
3	789	21	0.51	26
4	484	13	0.56	16
5	872	15	0.62	24
6	846	16	0.48	27

Equations (1) and (2): Allometric equations from [8] and [9]

$$log_{10}AGB_{standard} = -0.834 + 2.223(log_{10}DBH)$$

$$AGB_{musacea} = 0.030 * DBH^{2.13}$$

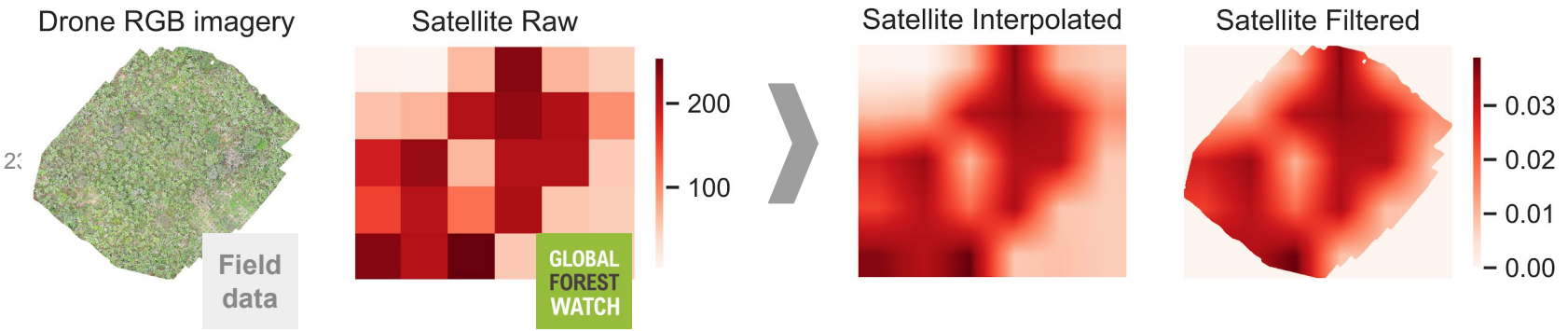
"Tackling the Overestimation of Forest Carbon with Deep Learning and Aerial Imagery"
Reiersen *et al.,* CCAI ICML 2021

Benchmarking satellite-based AGB density estimation against field data

Global Forest Watch product: *Aboveground live woody biomass density* [10]

- 30mx30m resolution, 70k GLAS observations with deep learning model
- Lidar-derived canopy metrics and region-specific allometric equations

Map interpolated and filtered on the locations for all field data sites:



"Tackling the Overestimation of Forest Carbon with Deep Learning and Aerial Imagery"
Reiersen *et al.*, CCAI ICML 2021

The satellite-based estimates significantly overestimates AGB density by a factor of 10

- The AGB density (kg/ha) per polygon was overestimated for **all of the 6 sites** with a factor ranging up to 10 times the field data

Site no.	Ground Truth	Filtered	Over estimation
1	19	176	×9.2
2	27	160	×5.9
3	24	47	×2.0
4	24	62	×2.6
5	17	19	×1.1
6	29	141	×4.9

Fig. 2 AGB density (kg/ha) of the field data (Ground truth) and of the satellite based estimations (Filtered)

"Tackling the Overestimation of Forest Carbon with Deep Learning and Aerial Imagery"
Reiersen *et al.*, CCAI ICML 2021

The critical role of IPLCs in co-designing AI

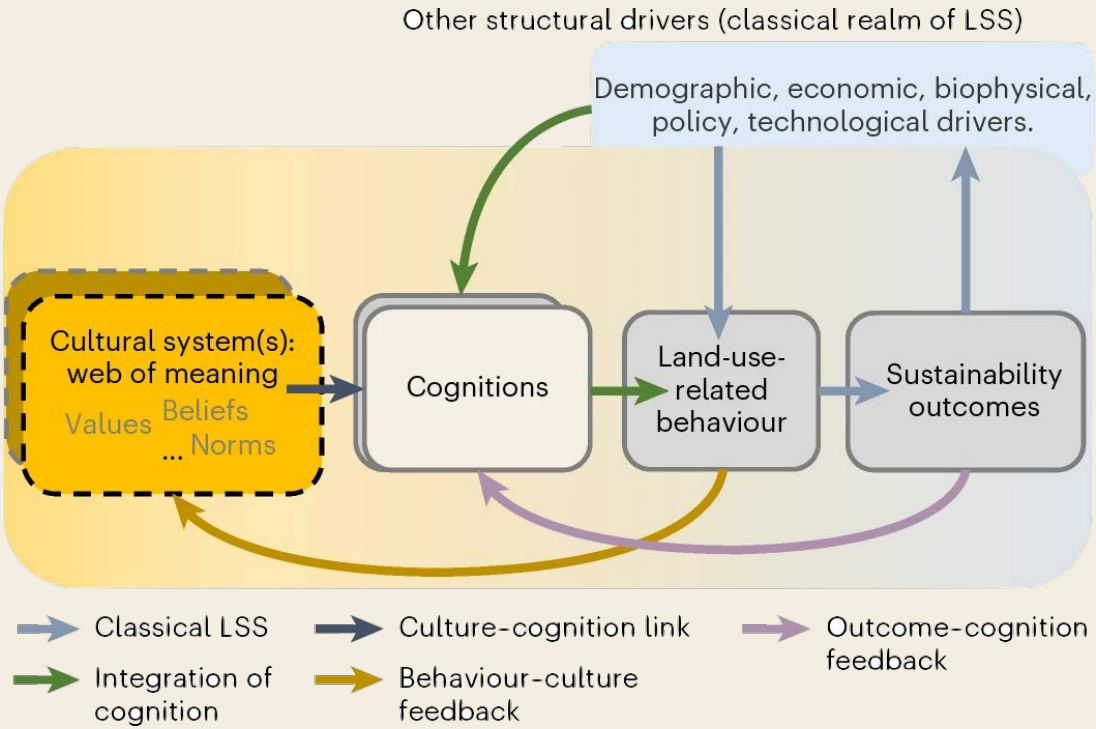
Important !

Indigenous People and Local Communities (IPLCs)

must be involved in the co-design of impactful technology

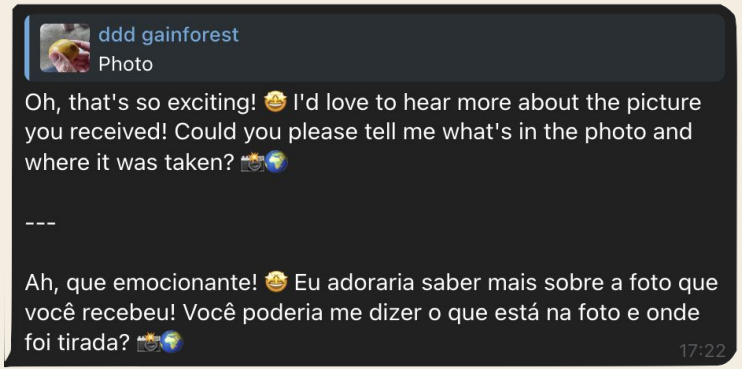
need to be one of the main beneficiaries of impactful technology

Culture and land systems



Hodel, L., le Polain de Waroux, Y. & Garrett, R.D. Characterizing culture's influence in land systems. Nat Sustain (2024). https://doi.org/10.1038/s41893-024-01381-z

Generative AI can provide an inclusive way for data collection



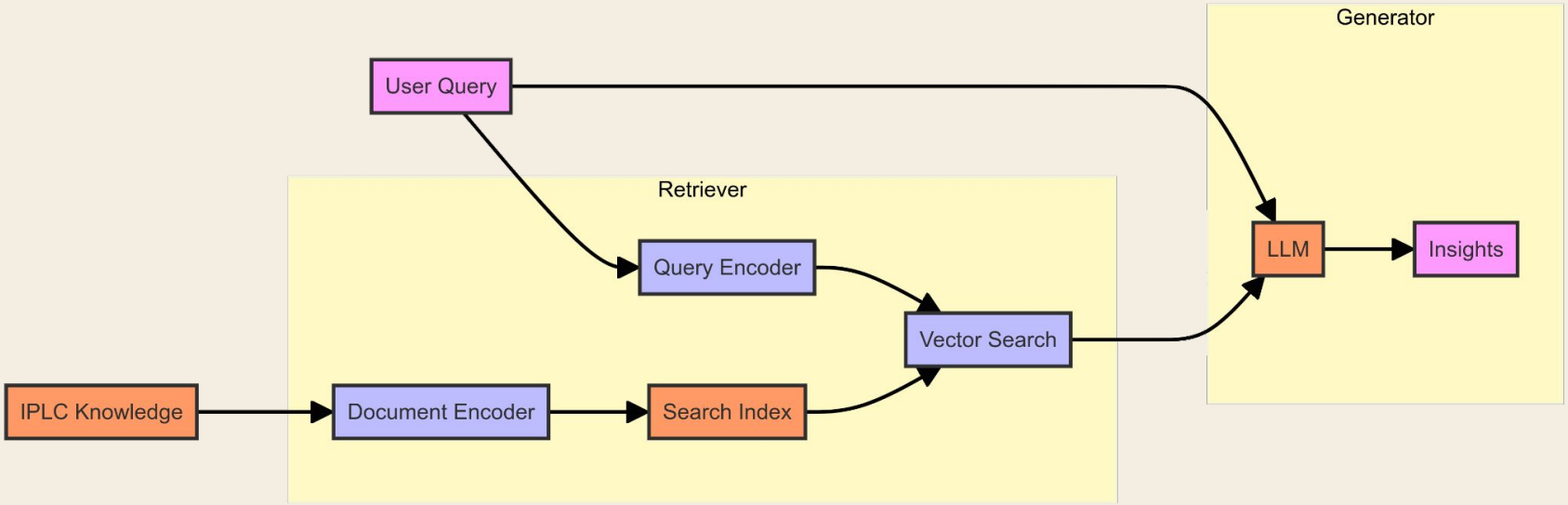
Test our Telegram Bot

Trust-building is critical

Trust through Data Sovereignty

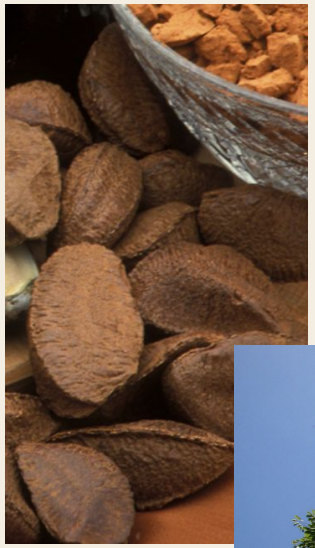
Trust through Ownership

RAG allows to integrate IPLC Knowledge

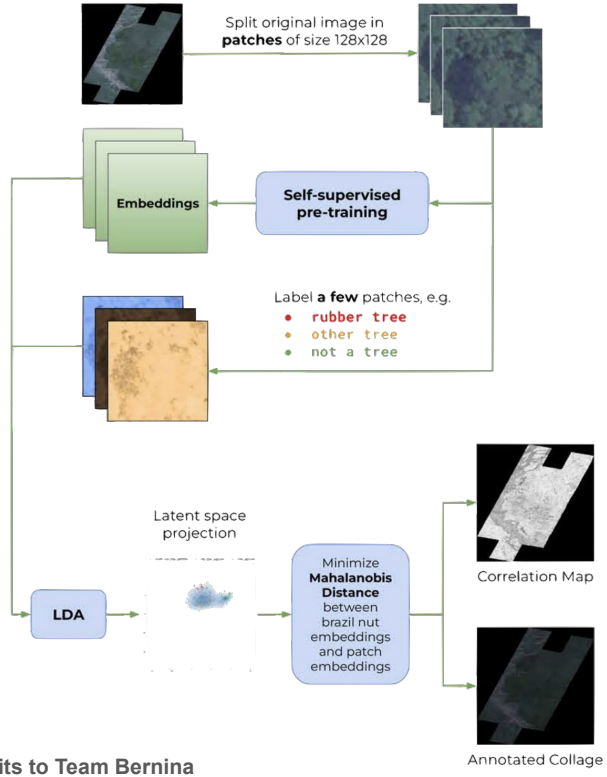


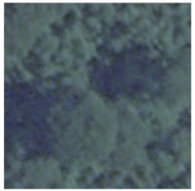
IPLC Knowledge as good priors

IPLC Knowledge as few shot examples

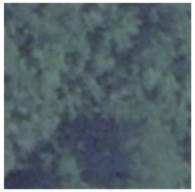


IPLC Knowledge as few shot examples

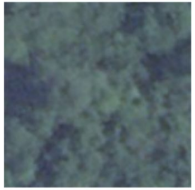




4_1245_6015

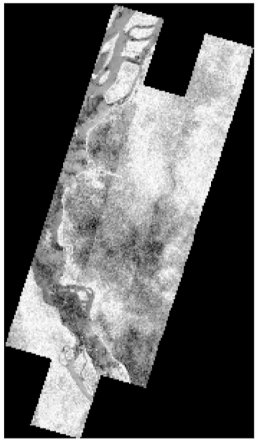


4_1266_6456



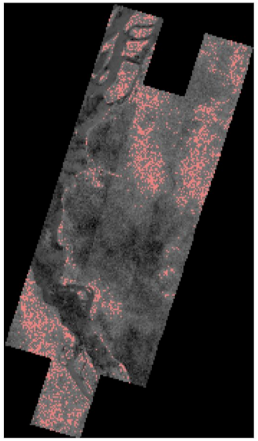
4_1308_6898

Correlation



Correlation map using LDA for dimensionality reduction and Mahalanobis distance as similarity measure.

Thresholding



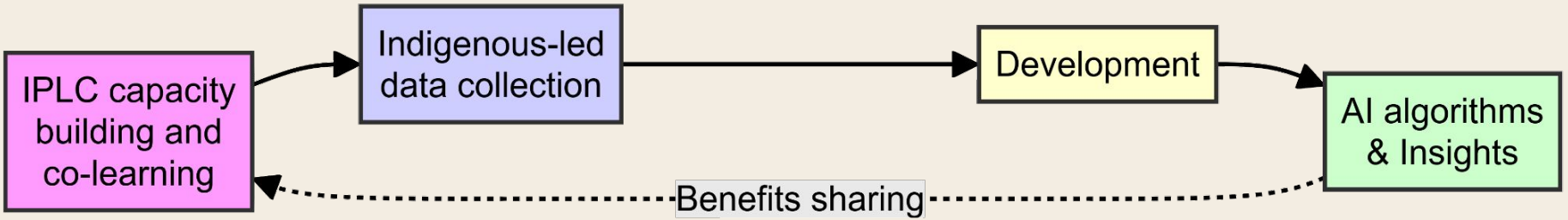
Threshold

0.95

0.50 1.00

Credits to Team Bernina

Co-designed AI needs to share benefits




This tool.

Lecture Overview

1. Climate Change: Forests and Land Systems
2. Saving Forests: Finance and Regulations
3. Artificial Intelligence: How can it help?
4. Equitable AI: The Role of IPLCs and Frontier Data
5. Future Work & Discussion

Lessons learned

- **Be transparent and open**
  - MRV algorithms and data influence policy and who receives payments or not, and can create dangerous feedback loops
- **Co-design with local communities**
  - Data ownership
  - Scenario-specific metrics
  - Goodhart's Law: "When a measure becomes a target, it ceases to be a good measure"
- **Program Benefit-Sharing In**
  - FAIR (findable, accessible, interoperable and reusable) vs Benefit-Sharing (CARE)
  - Monetary and Non-Monetary Benefits


INTHA

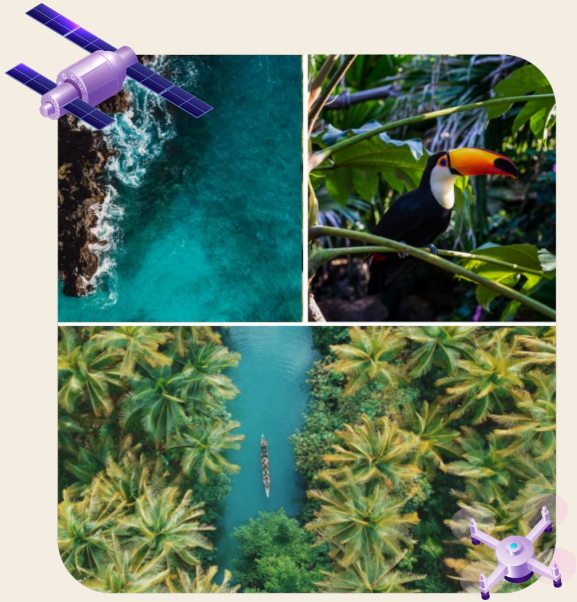

Number of citizen scientists that identified at least one species: 30

Thank you!

David Dao

@dwddao

@gainforestnow



Appendix

More research, etc …

Challenge

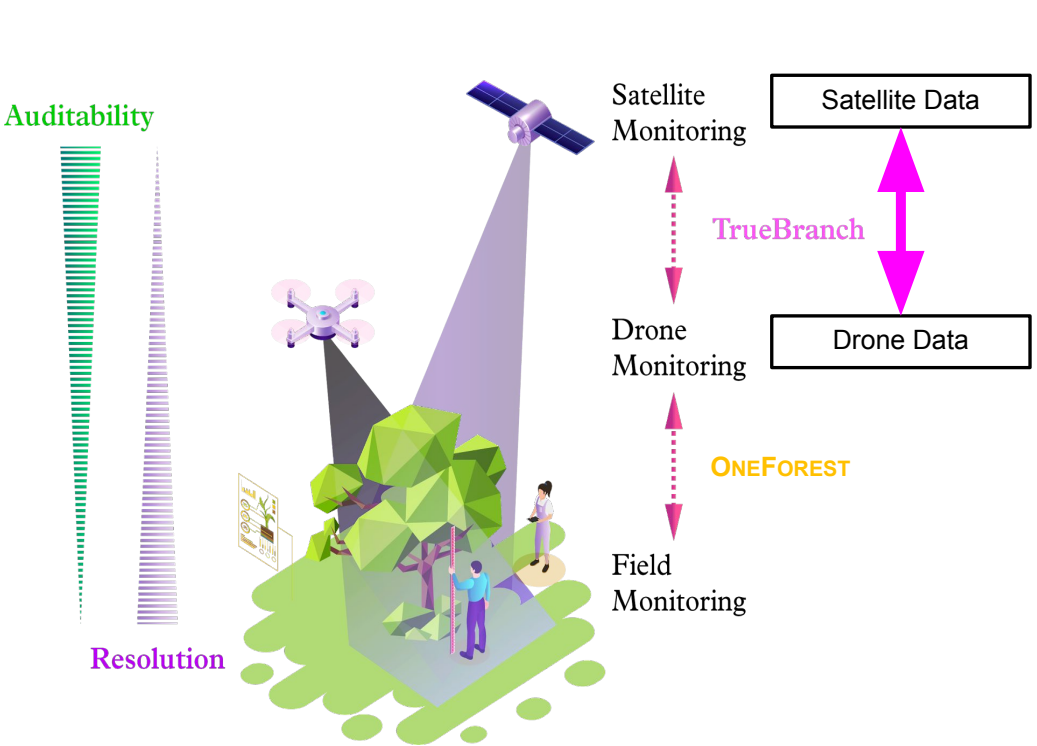
Automated forest validation opens up possibility of *untruthfully* reported imagery

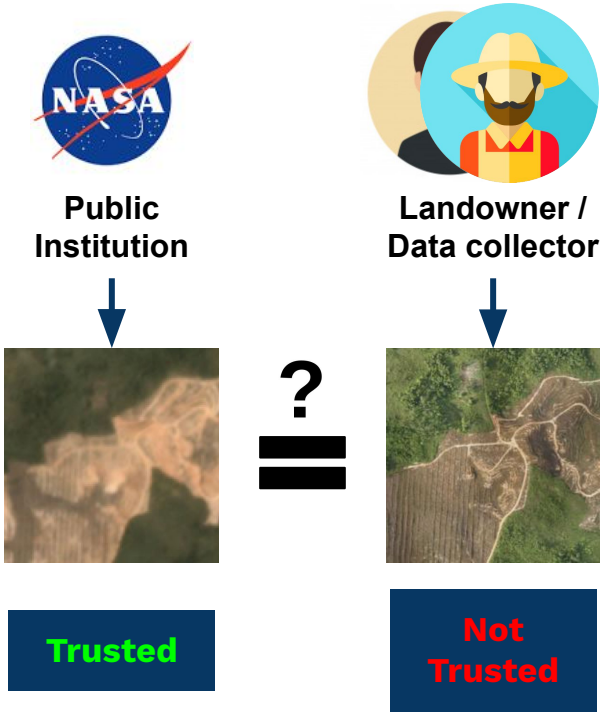
Attack vectors

Reported Land-Use				
	true time **true** location	**wrong** time **true** location	**true** time **wrong** location	**modified image**
Detected Forest Cover	**medium**	**high**	**high**	**high**

Simona Santamaria*, David Dao*, Björn Lütjens*, TrueBranch: Metric Learning-Based Verification of Forest Conservation Projects, 2020 ICML Workshop (CCML), *equal contribution, https://arxiv.org/abs/2004.09725

Novel opportunities through data fusion

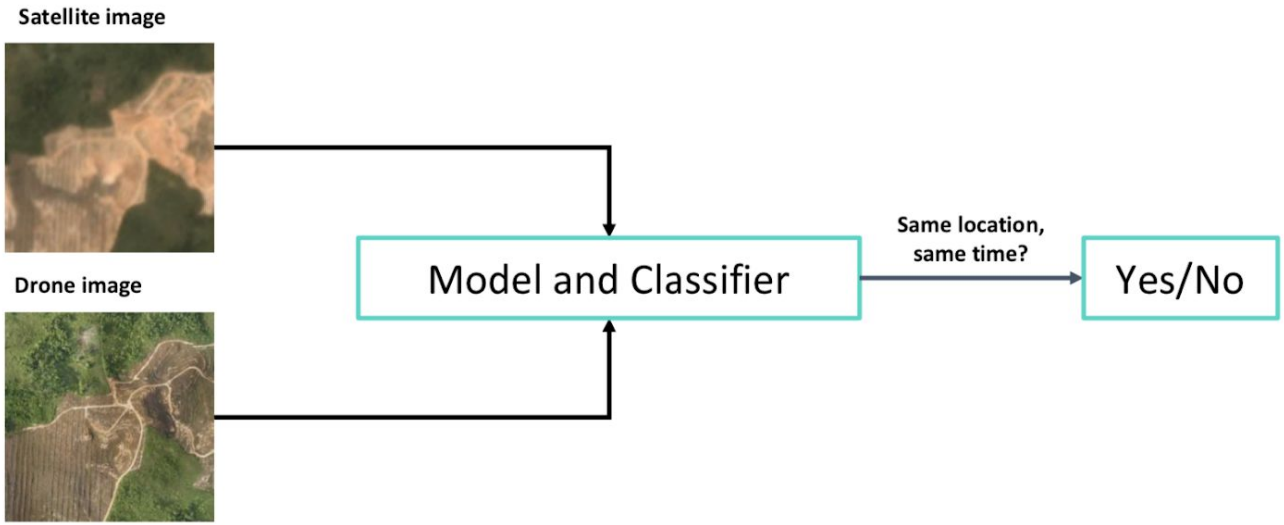




Classifying truthfulness

How to distinguish truthful imagery from untruthful imagery?

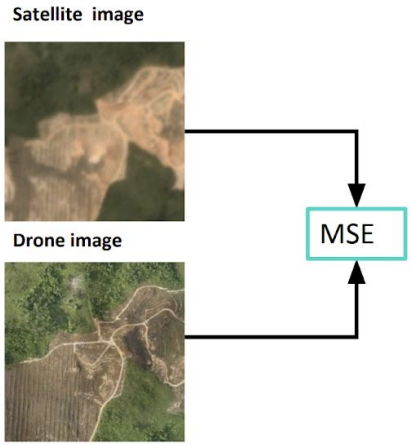
- Image Registration: Matching Drone images with Satellite images



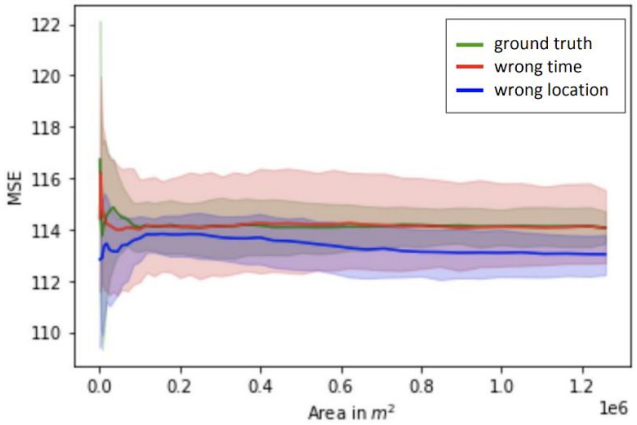
Simona Santamaria*, David Dao*, Björn Lütjens*, TrueBranch: Metric Learning-Based Verification of Forest Conservation Projects, 2020 ICML Workshop (CCML), *equal contribution, https://arxiv.org/abs/2004.09725

Classifying truthfulness

- Nominal distance metrics of MSE in pixels space



$$MSE = \frac{1}{mn} \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} [A(i,j) - B(i,j)]^2$$

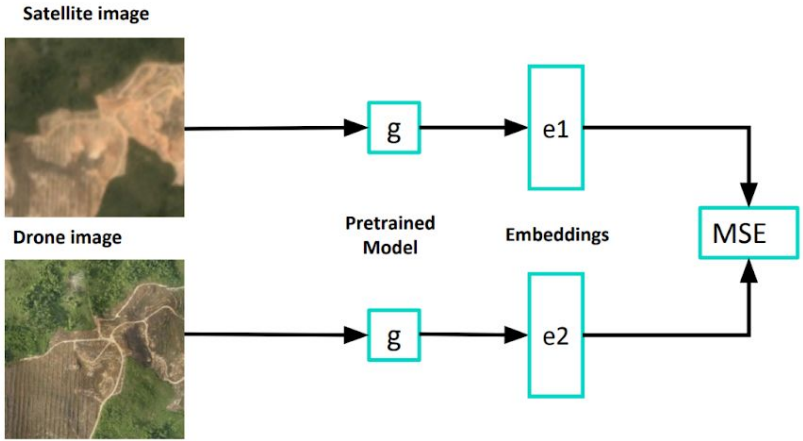


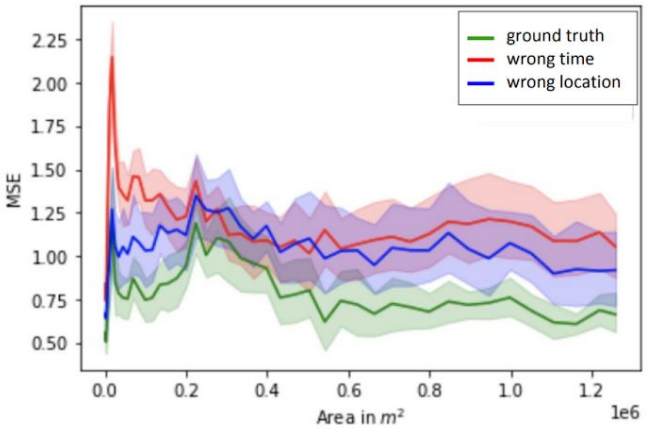
Simona Santamaria*, David Dao*, Björn Lütjens*, TrueBranch: Metric Learning-Based Verification of Forest Conservation Projects, 2020 ICML Workshop (CCML), *equal contribution, https://arxiv.org/abs/2004.09725

Classifying truthfulness

- Nominal distance metrics of MSE in feature space

$$MSE = \frac{1}{mn} \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} [A(i,j) - B(i,j)]^2$$

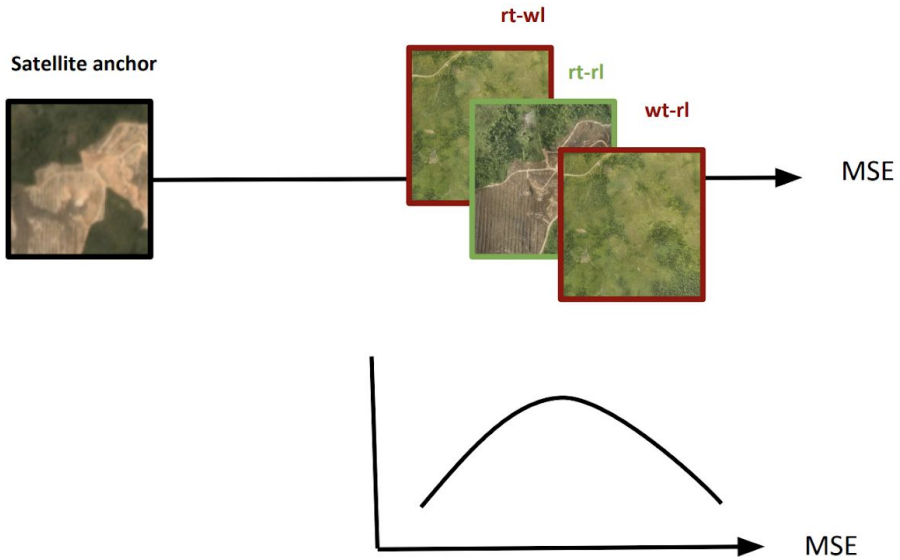




Simona Santamaria*, David Dao*, Björn Lütjens*, TrueBranch: Metric Learning-Based Verification of Forest Conservation Projects, 2020 ICML Workshop (CCML), *equal contribution, https://arxiv.org/abs/2004.09725

Classifying truthfulness

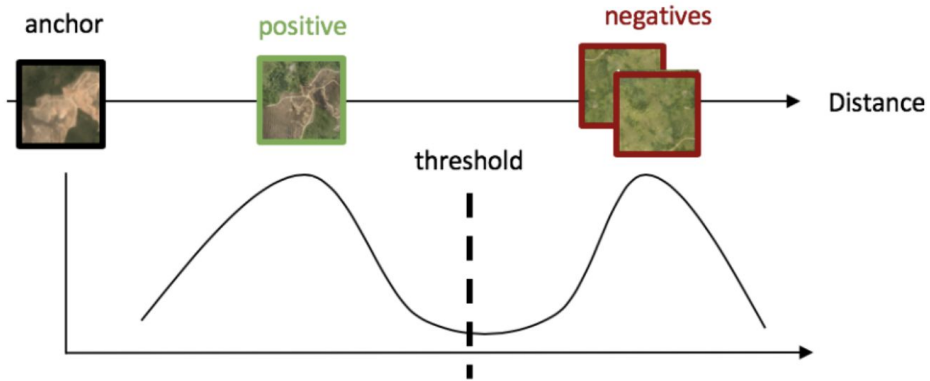
- MSE in pixel space and RESISC-45 feature space not sufficient



Simona Santamaria*, David Dao*, Björn Lütjens*, TrueBranch: Metric Learning-Based Verification of Forest Conservation Projects, 2020 ICML Workshop (CCML), *equal contribution, https://arxiv.org/abs/2004.09725

Metric Learning

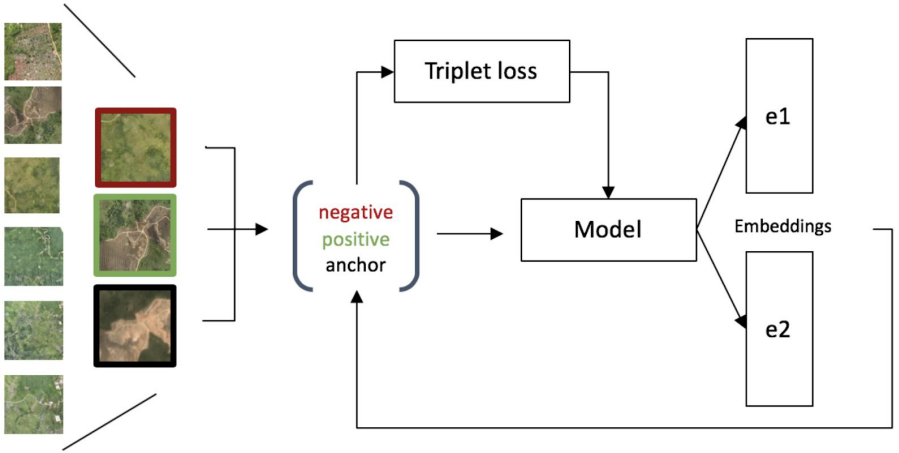
The distance between the anchor and positive image is decreased while the distance between the anchor and negative image is increased.

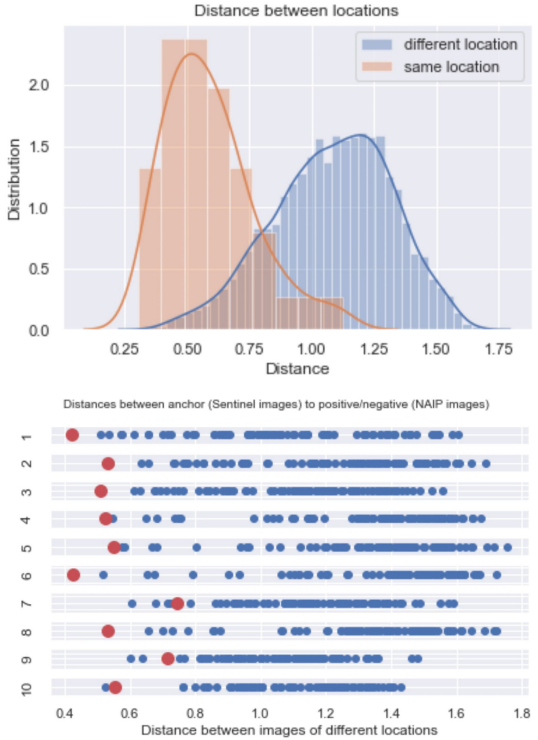


$$L(a, p, n) = max(|f(a) - f(p)|^2 - |f(a) - f(n)|^2 + \alpha, 0)$$

Simona Santamaria*, David Dao*, Björn Lütjens*, TrueBranch: Metric Learning-Based Verification of Forest Conservation Projects, 2020 ICML Workshop (CCML), *equal contribution, https://arxiv.org/abs/2004.09725

TrueBranch: Metric Learning-based Verification





TrueBranch enables the verification of truthfully reported drone imagery from untrusted parties

Simona Santamaria*, David Dao*, Björn Lütjens*, TrueBranch: Metric Learning-Based Verification of Forest Conservation Projects, 2020 ICML Workshop (CCML), *equal contribution, https://arxiv.org/abs/2004.09725

Third-Party Monitoring, Reporting and Verification (MRV) is manual and expensive

Tree metrics e.g. Diameter at Breast Height (DBH), species, etc

Allometric equations

Aboveground biomass (AGB) kg

Biomass = 2x Carbon

Carbon stock kg of CO2

300 USD/Ha

for forest monitoring, verification and reporting

Source: FAO

Results-based payments require monitoring, reporting and verification (MRV)

$25bn →  → 

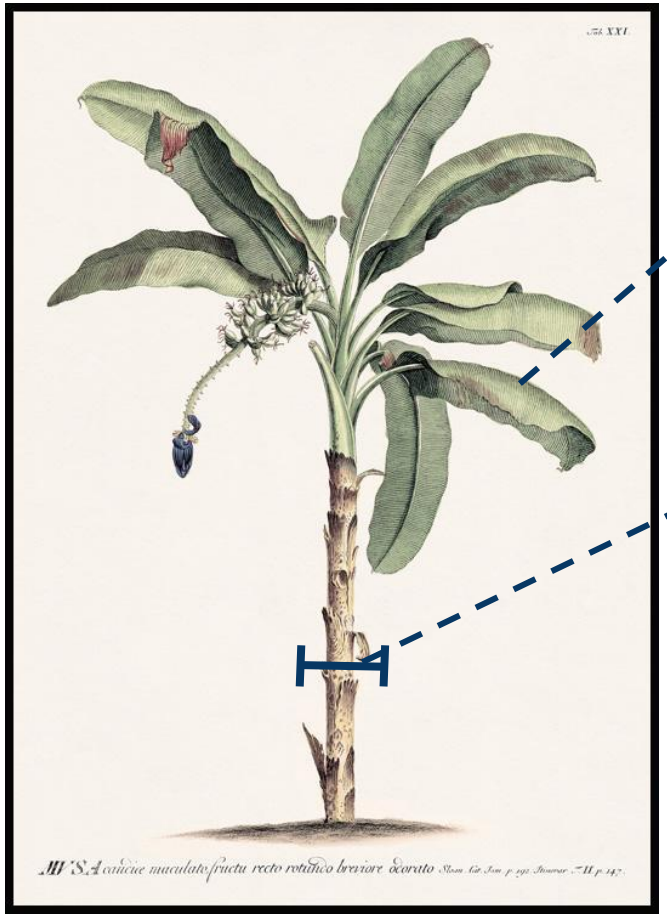
Public funds reserved for results-based payments in forestry

But overlapping land claims, slow performance measurement and missing trust

Make funds inaccessible to local communities

Source: REDD+

Verification?



Species: **Musaceae**

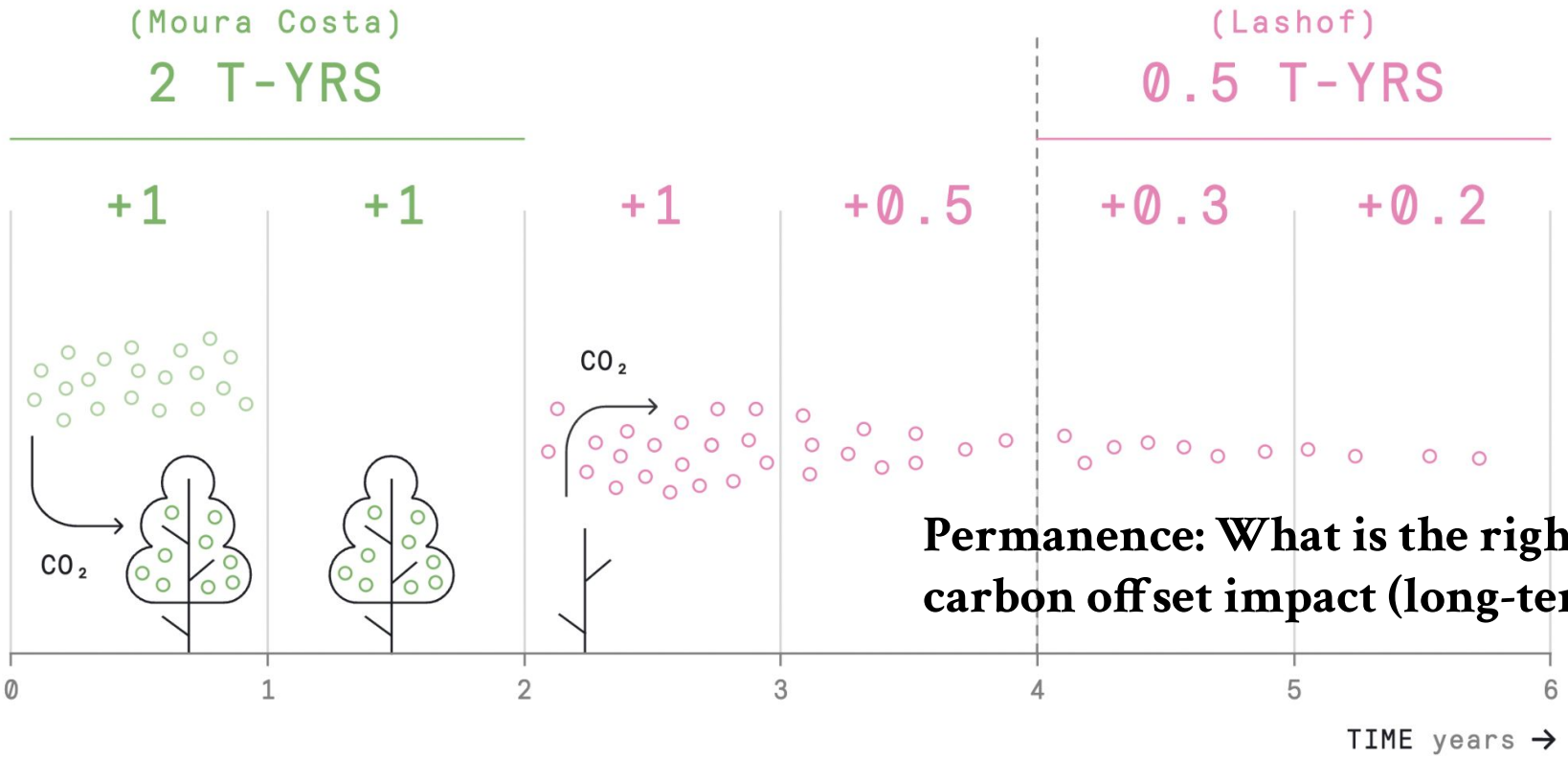
$$AGB_{musacea} = 0.030 * DBH^{2.13}$$

Diameter at Breast Height (DBH) : **0.21 m**

$$AGB_{musacea} = 0.030 * \boxed{0.21}^{2.13}$$

Biomass: **0.001 kg/m2**

Permanence?



Permanence: What is the right carbon offset impact (long-term)?

Baseline?

<u>Baseline</u>: What is the right baseline to issue credits? What would have happened without the funding?