Enhancing Biomass and Carbon Stock Estimation in Tanzanian Forests: Integrating Earth Observation and Machine Learning for Sustainable Forest Management and Food Security (TANZEO-BioStock)

PIs: Sadegh Jamali, Ernest William Mauya
Team members: Torbern Tagesson, Filbert Francis, Soheil Zaghian, Seyedehsan Khankeshizadeh, Justo Jonas, Sami Madundo, Diana Tesha

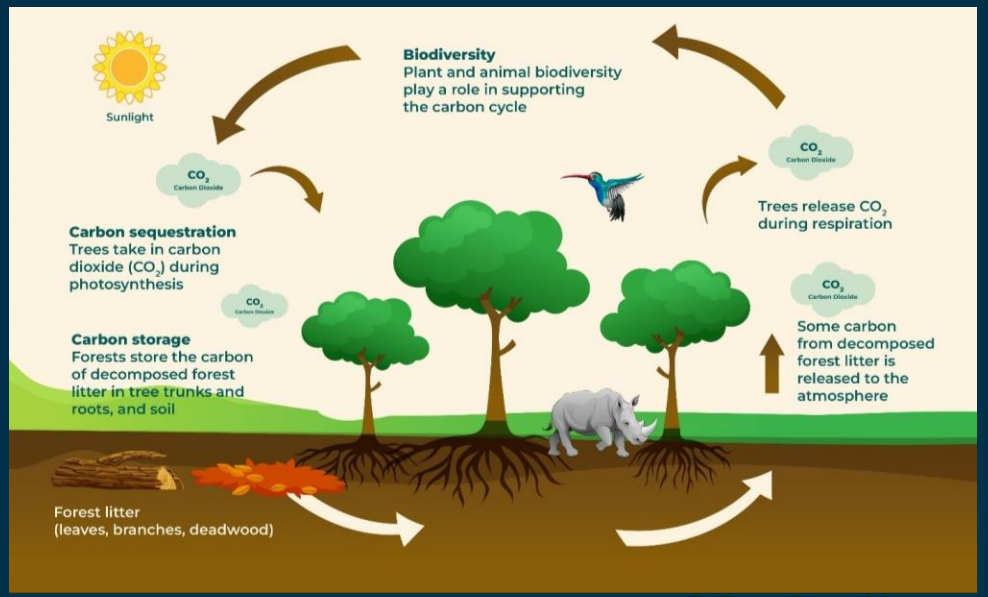
Scientific Background

Forest ecosystems play a critical role in the global carbon cycle, acting as major carbon sinks and supporting climate regulation.

Key structural variables

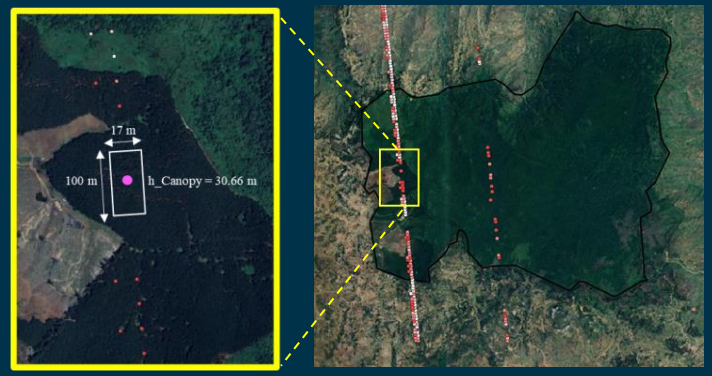
- **Forest Canopy Height (FCH)**
- **Above-Ground Biomass (AGB)**
- **Above-Ground Carbon (AGC)**

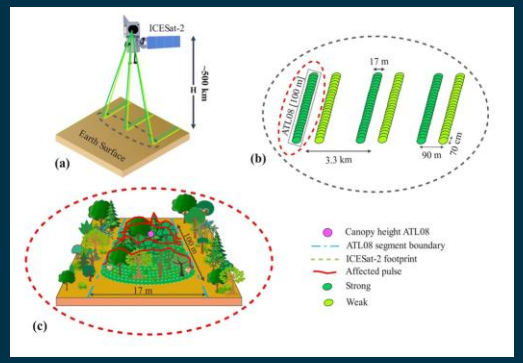
  are essential for understanding forest structure, ecosystem health, and carbon storage dynamics.

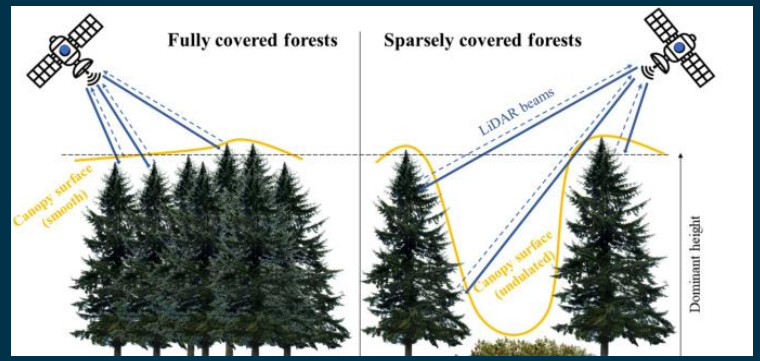


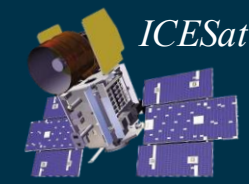
Scientific Background

Spaceborne LiDAR missions (e.g., ICESat-2, GEDI) provide highly accurate vertical forest structure measurements.









However, these datasets are spatially sparse (along-track sampling) and cannot directly produce continuous maps.

Multi-source remote sensing data

- **Sentinel-2 (optical)** → vegetation properties
- **Sentinel-1 (SAR)** → structural information
- **DEM** → topographic influence

can be fused with LiDAR data to generate **wall-to-wall forest structure maps**.

Overall Goal and Objectives

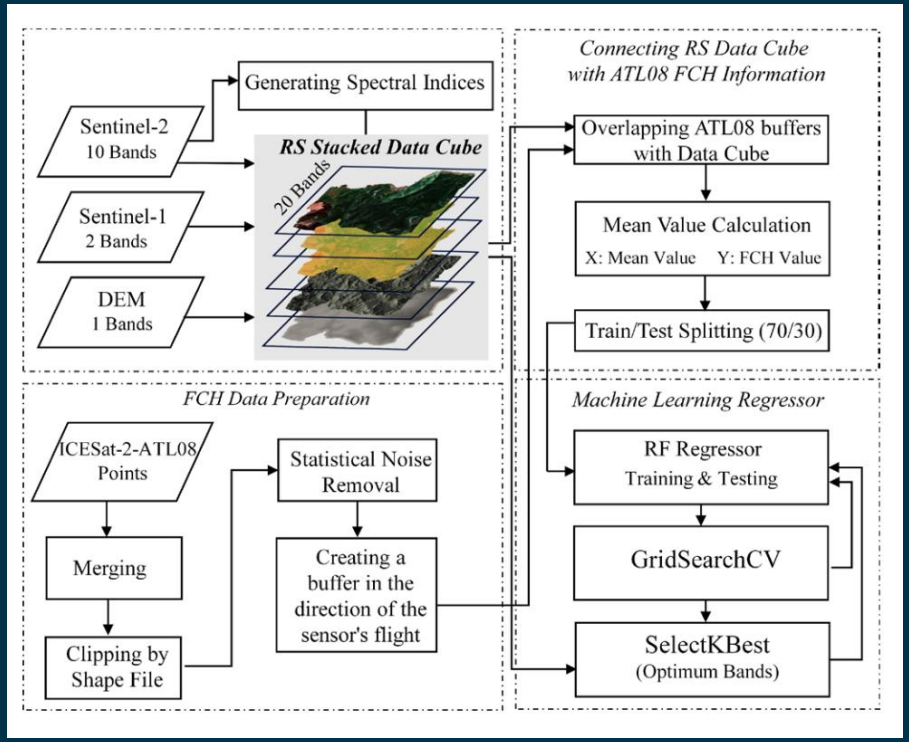
Goal

Develop a scalable and reproducible framework for mapping forest structure, biomass, and carbon stocks in tropical montane forests using Earth Observation (EO) data and machine learning.

Objectives

Phase 1 – ICESat-2-based FCH Mapping

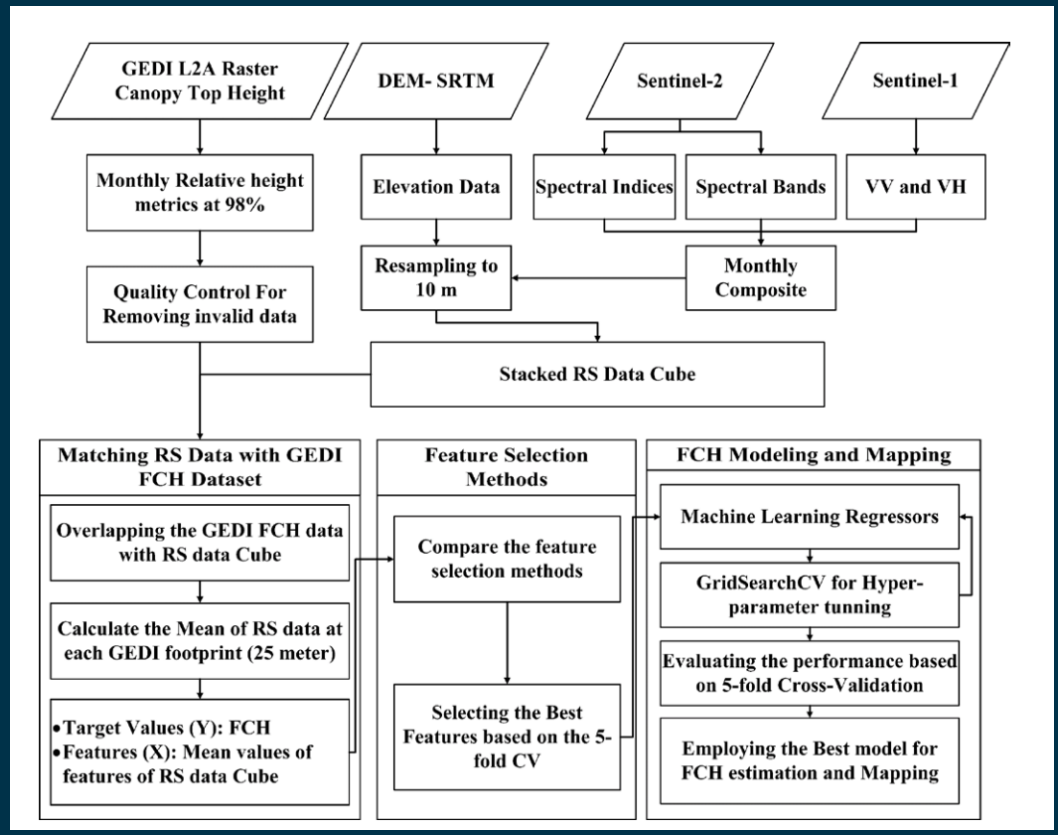
- Generate **wall-to-wall canopy height maps** using ICESat-2 ATL08 data
- Integrate multi-source satellite data with machine learning (Random Forest)
- Address spatial discontinuity of LiDAR observations



Overall Goal and Objectives

Phase 2 – GEDI-based FCH Mapping

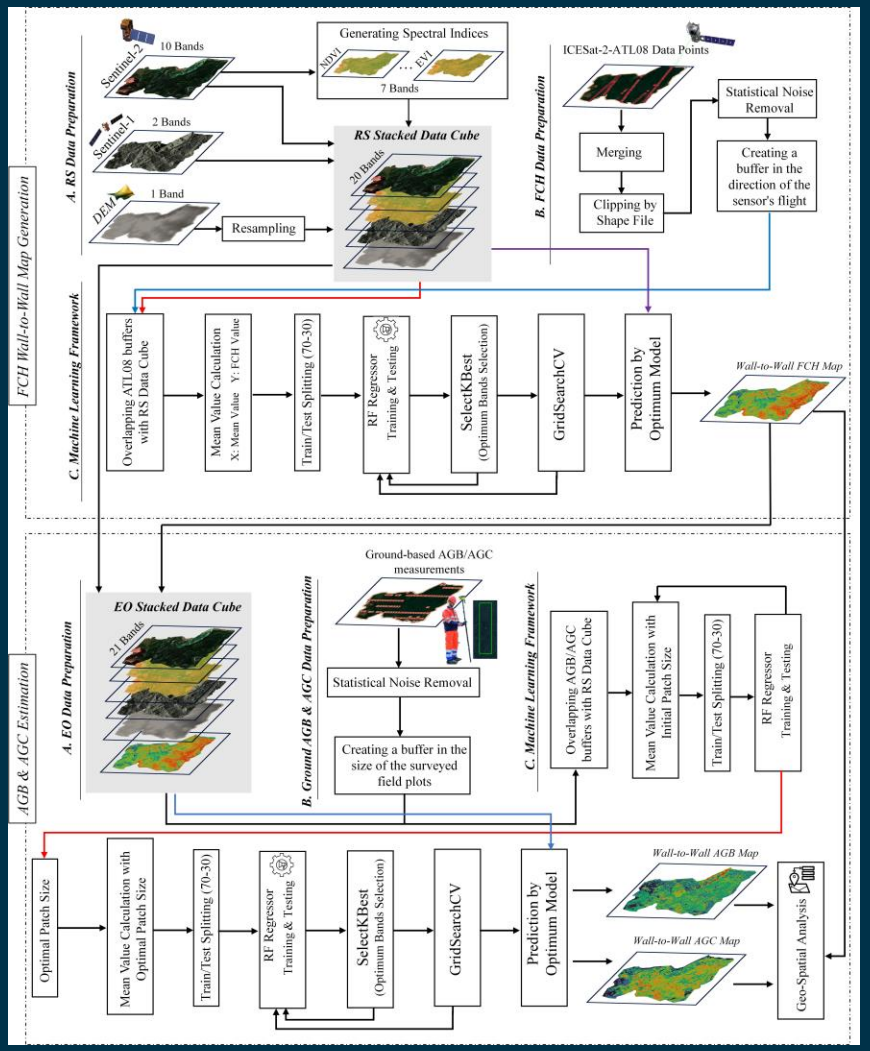
- Improve canopy height estimation using GEDI L2A data
- Evaluate feature selection methods and ML models
- Identify optimal predictors for FCH mapping



Overall Goal and Objectives

Phase 3 – Biomass, Carbon & Risk Mapping

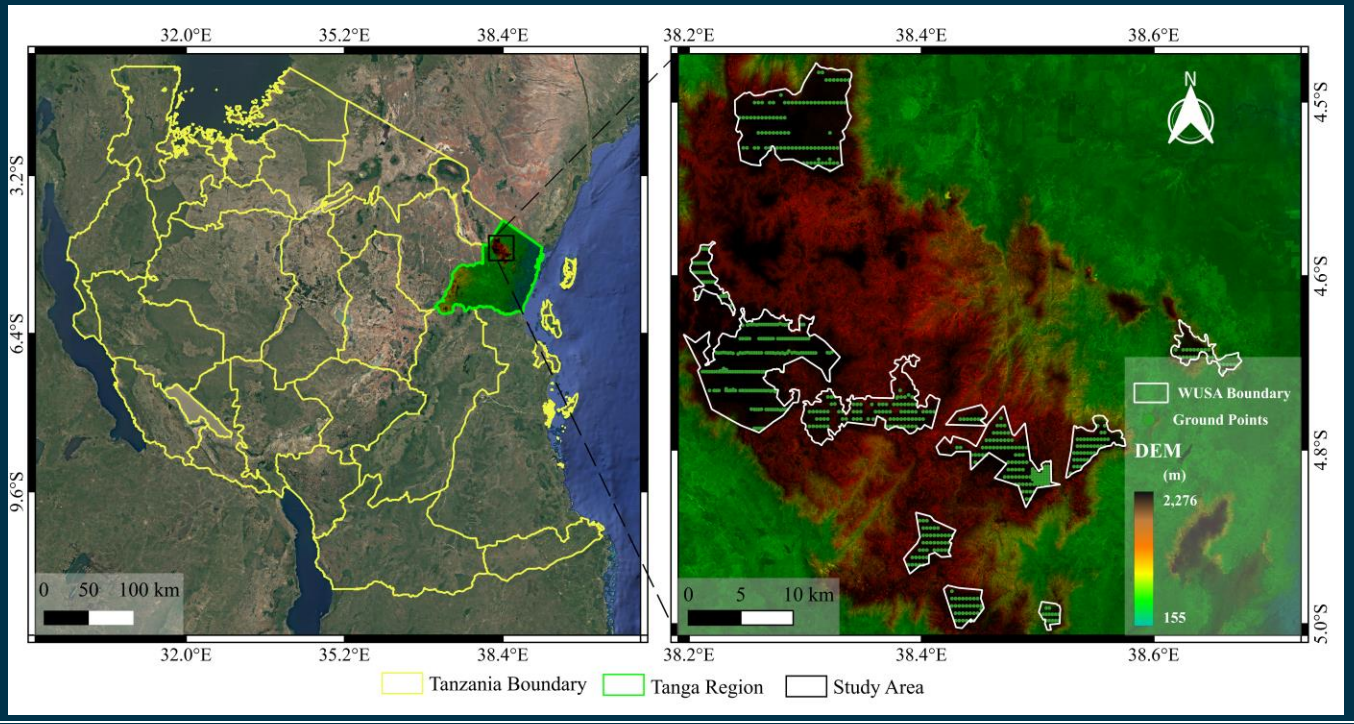
- Use wall-to-wall FCH as a structural predictor for: AGB and AGC
- Develop uncertainty-aware models
- Map forest degradation risk and restoration potential



Study Area

- West Usambara Mountains (WUSA), Tanzania
- Part of the Eastern Arc Mountains – a global biodiversity hotspot
- Complex terrain (400–2200 m elevation) with high structural heterogeneity
- Ideal test site for forest structure and biomass modeling

Usambara Mountain Forest (tropical rainforests, district covering an area of 3,297 km2 with an estimated population of 350,959)



Research Outline

- **Phase 1: ICESat-2-based FCH Mapping**

Goal: Generate continuous canopy height map

Steps:

- Multi-source data preparation:
  1. Sentinel-2 (spectral bands + vegetation indices)
  2. Sentinel-1 (VV, VH backscatter)
  3. DEM (topography)

- ICESat-2 ATL08 preprocessing:
  1. Noise filtering and quality control
  2. Creation of footprint buffers (~17m × 100m)

- Data integration:
  1. Extract RS features within each footprint

- Machine Learning:
  1. Random Forest regression
  2. Feature selection + hyperparameter tuning

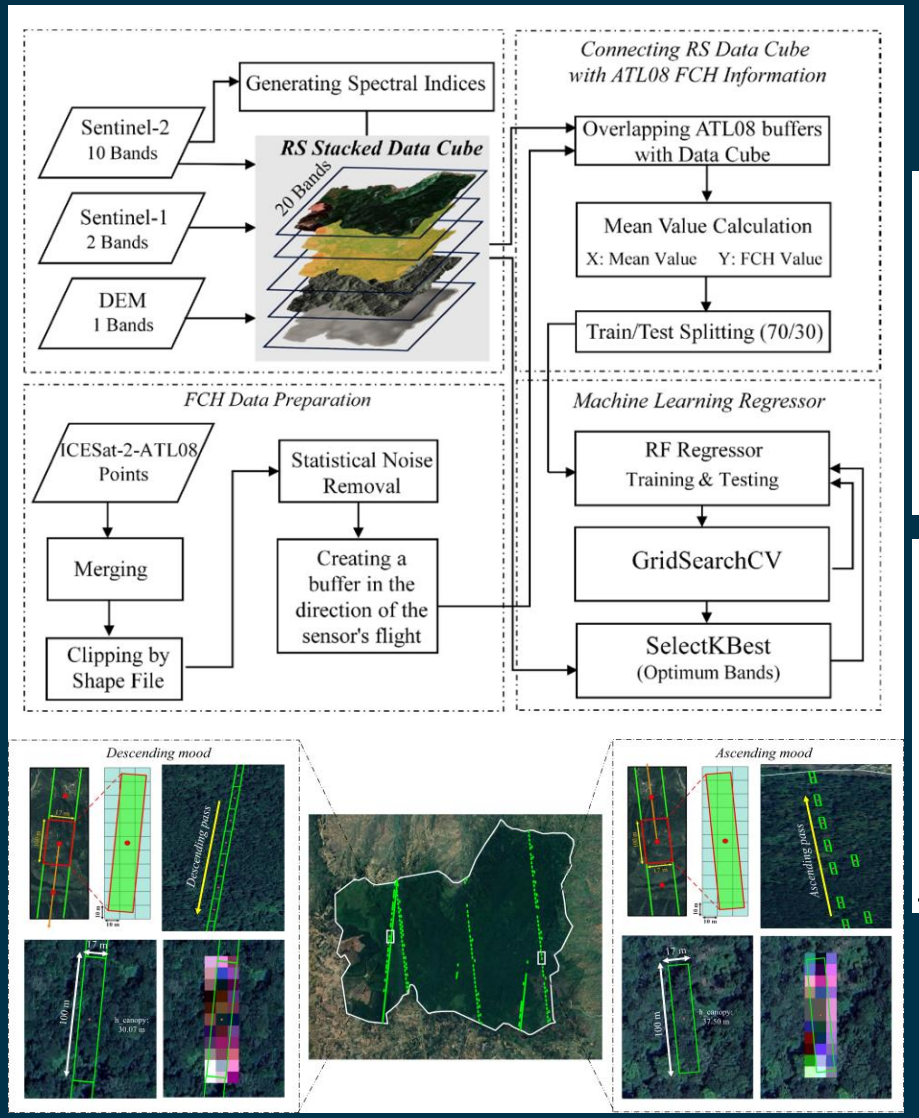


TABLE II. LIST OF DATASETS USED

Name	Acquisition time	Spatial resolution	Temporal resolution
ICESat-2	2024-10-13 (Desc) /2024-11-03 (Asc) /2024-11-11 (Desc)	-	91 days
Sentinel-2	Median Image (2024-10-01:2024-11-30)	10 m	5 days
Sentinel-1	Median Image (2024-10-01:2024-11-30)	10 m	6 days
DEM	2024-09-21	30 m	-

TABLE. III. SUMMARY OF THE FEATURES USED

Data source	Utilized features	Description
Sentinel-2	Spectral Bands	B2 (Blue [490 nm])
		B3 (Green [560 nm])
		B4 (Red [665 nm])
		B5 (Red Edge 1 [705 nm])
		B6 (Red Edge 2 [749 nm])
		B7 (Red Edge 3 [783 nm])
		B8 (NIR1 [842 nm])
		B8A (NIR2 [865 nm])
		B11 (SWIR1 [1610 nm])
		B12 (SWIR2 [2190 nm])
Sentinel-1	Polarizations in Asc and Desc orbits	VV
		VH
DEM-SRTM	Topographic information	Elevation

Research Outline

- **Phase 2: GEDI-based FCH Mapping**

Goal: Improve canopy height estimation using multimodal data

Steps:

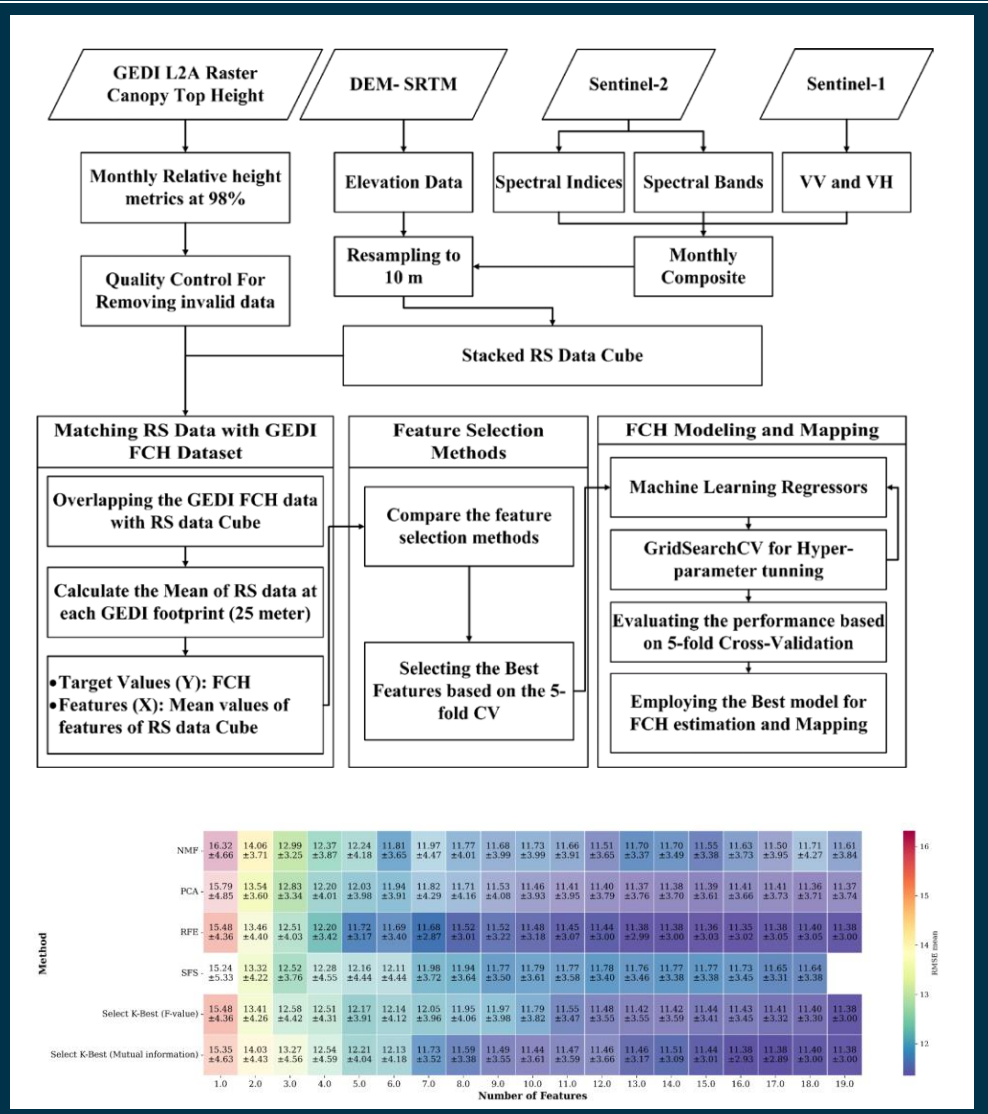
- Data harmonization:
  1. GEDI L2A canopy height (rh98)
  2. Monthly composites of Sentinel-1 & Sentinel-2

- Feature engineering:
  1. Spectral bands + indices (NDVI, NDMI, EVI, etc.)

- Feature selection:
  1. RFE, PCA, SelectKBest, SFS

- Model comparison:
  1. RF, SVR, GBR, OLS

- Best model selection:
  1. Based on cross-validation (RMSE, R^2, MAE)



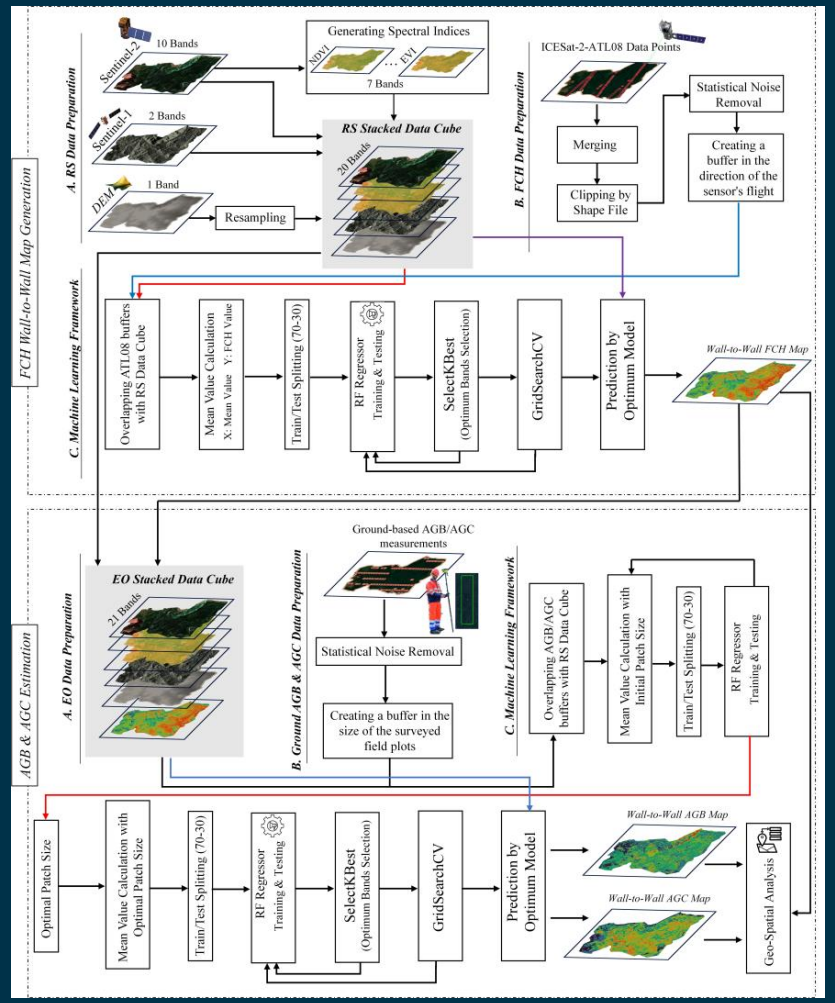
Research Outline

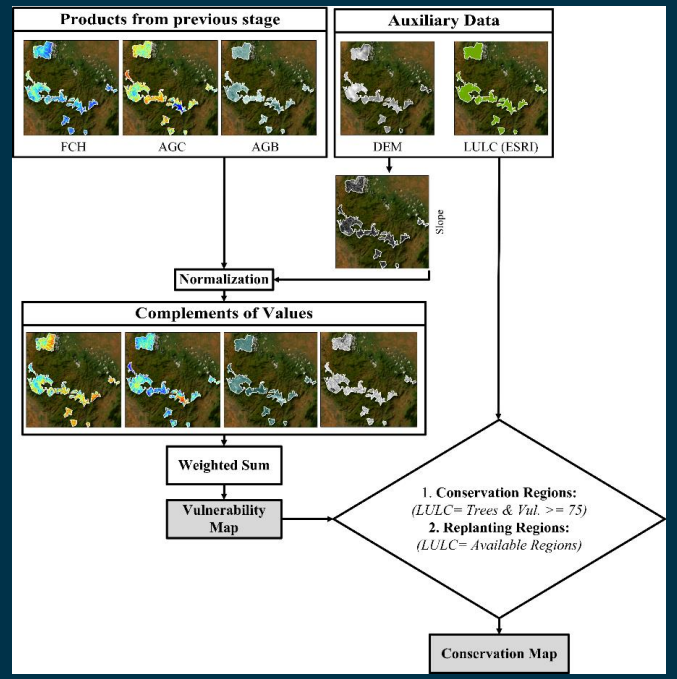
- **Phase 3: Biomass, Carbon & Risk Mapping**

Goal: Estimate forest carbon dynamics and vulnerability

Steps:

- Use FCH map as structural predictor
- Integrate:
  1. RS data cube + FCH
  2. Field inventory data (AGB/AGC)
- Machine Learning:
  1. Random Forest regression for AGB & AGC
- Uncertainty analysis:
  1. Conformal prediction (MAPIE)
  2. Monte Carlo simulations
- Vulnerability mapping:
  1. Indicators: FCH, AGB, AGC, slope
  2. Fuzzy AHP + PCA weighting





Key Innovation

- Multi-sensor data fusion (LiDAR + Optical + SAR + DEM)
- Wall-to-wall forest structure mapping at 10 m resolution
- Integration of **structure → biomass → carbon → risk**
- Uncertainty-aware modeling framework

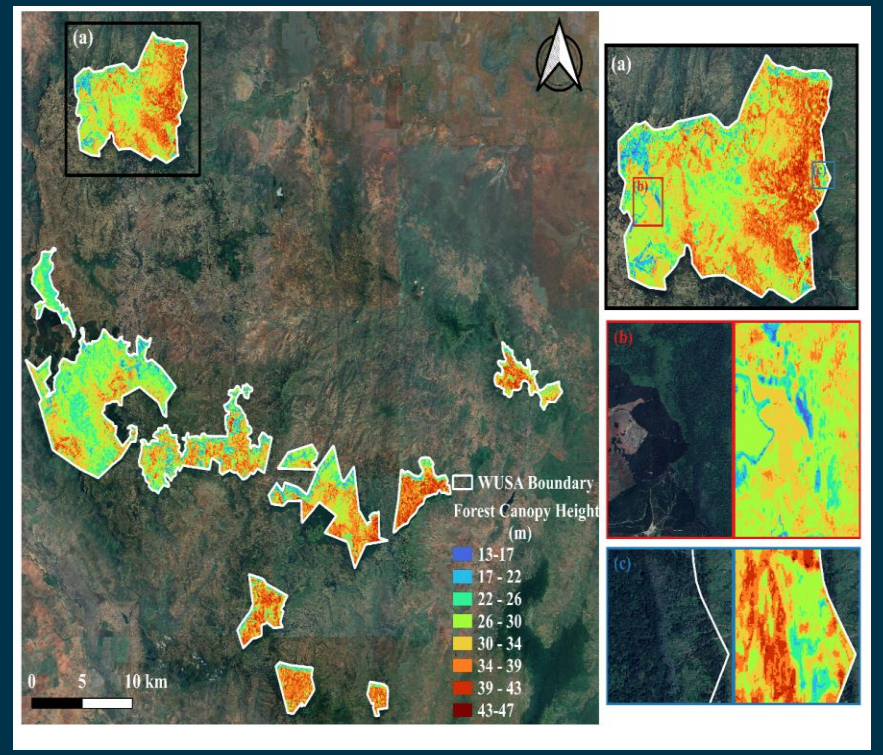
Results

Phase 1: ICESat-2-based FCH Mapping

- Successfully generated a **wall-to-wall Forest Canopy Height (FCH) map** from sparse ATL08 observations
- Integration of multi-source data enabled spatial continuity
- Random Forest model achieved:
  - **$R^2 \approx 0.29$**
  - **RMSE ≈ 8.6 m**
- Key predictors:
  - Topographic variables (DEM)
  - Spectral variables (SWIR2, RedEdge1)
  - SAR backscatter (VV)

TABLE. IV. QUANTITATIVE PERFORMANCE COMPARISON OF ML REGRESSOR MODELS

ML models	RMSE	MAE	R^2 Score
SVR	9.9160	7.7014	0.0598
RFR	8.8368	6.7038	0.2533
RFR+SelectKBest	8.6753	6.5672	0.2803
RFR+SelectKBest+GridSearch	8.6406	6.6482	0.2861



Results

Phase 2: GEDI-based FCH Mapping

•Improved canopy height estimation using **GEDI L2A data + multimodal RS data**

•Feature selection significantly enhanced model performance

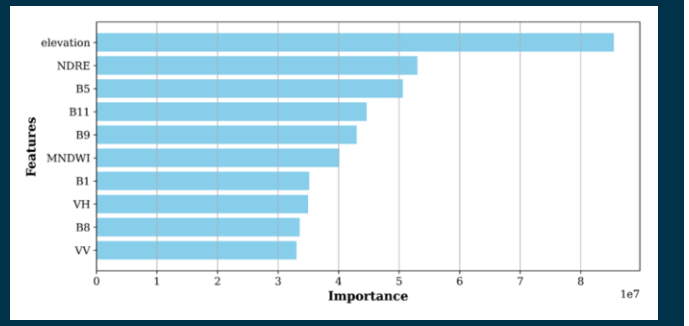
Key Findings:

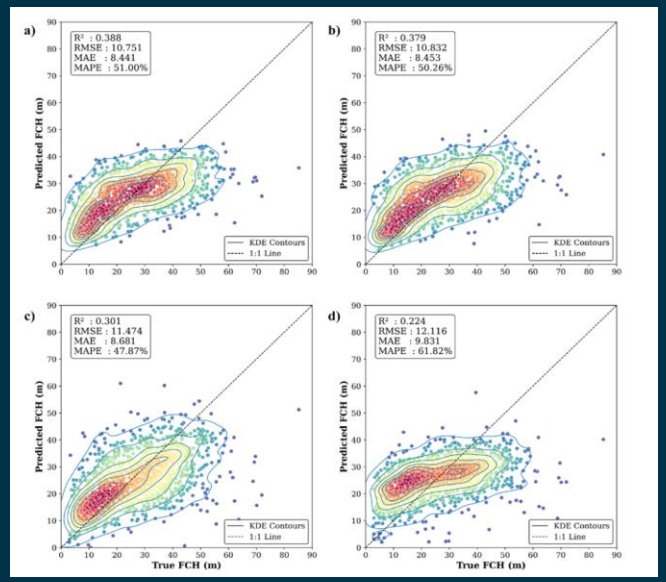
•Best feature set selected using **RFE (10 features)**

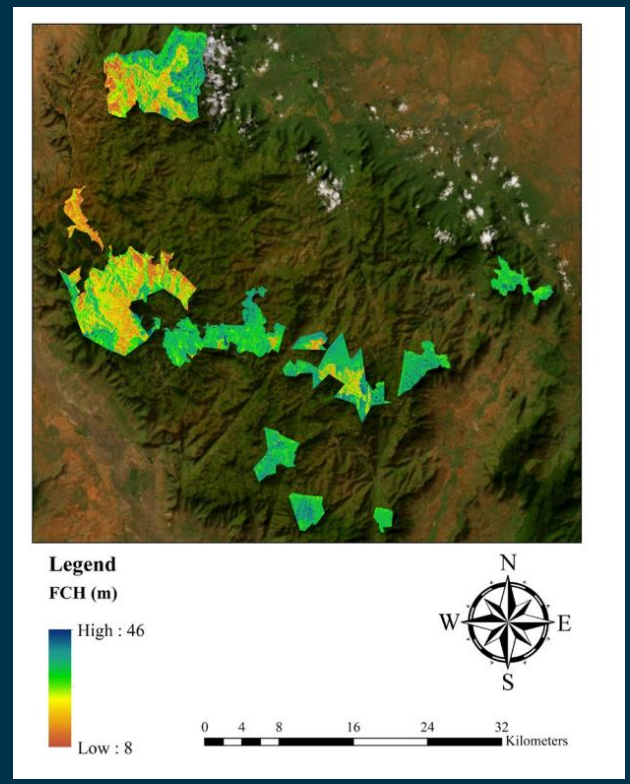
•Random Forest outperformed other models:
- Better accuracy than SVR, GBR, and OLS

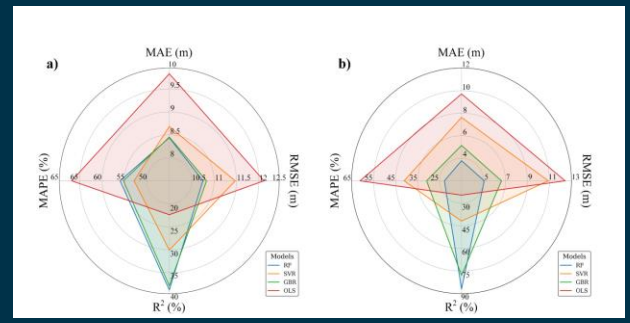
•Optimal performance:
- **RMSE ≈ 10.87 ± 1.27 m**
- Stable and reliable predictions

•Important variables:
- Sentinel-2 spectral bands
- Topography (DEM)
- SAR + vegetation indices









Results

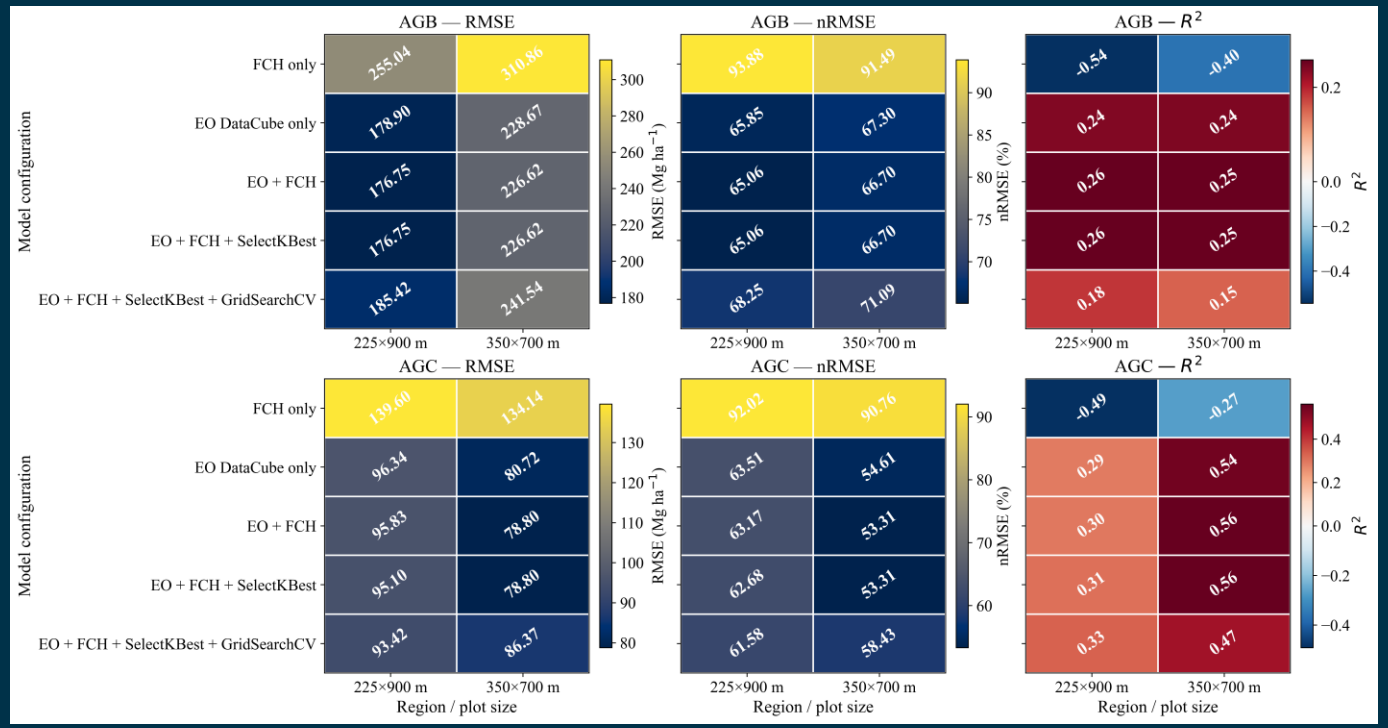
Phase 3: Biomass & Carbon Estimation

Biomass (AGB) & Carbon (AGC)

- Integration of **FCH + RS data** significantly improved estimation accuracy
- Best performance achieved using combined predictors (RS + FCH)

AGB Results:
- $R^2 \approx 0.25–0.26$
- Lower RMSE compared to single-source models

AGC Results:
- R^2 up to **0.56** (best performance)



Results

Wall-to-Wall Mapping Outputs

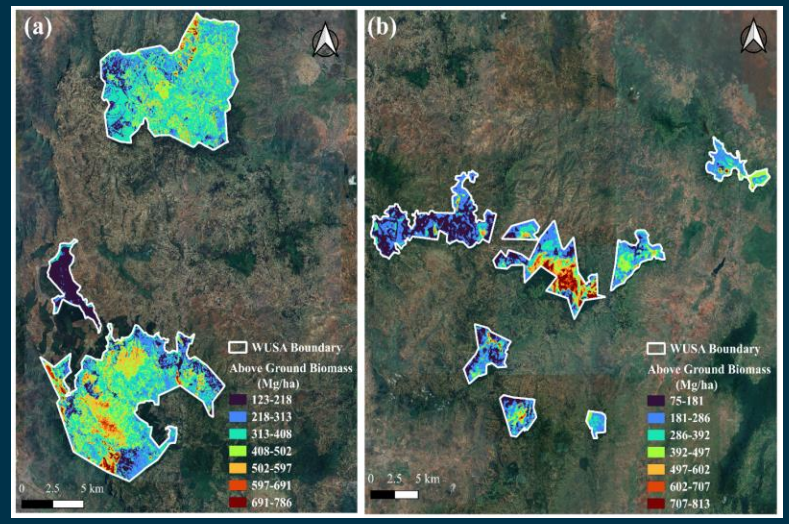
- Produced high-resolution (**10 m**) maps of:
  - Forest canopy height (FCH)
  - Above-ground biomass (AGB)
  - Above-ground carbon (AGC)

- Captured **strong spatial heterogeneity** across the study area

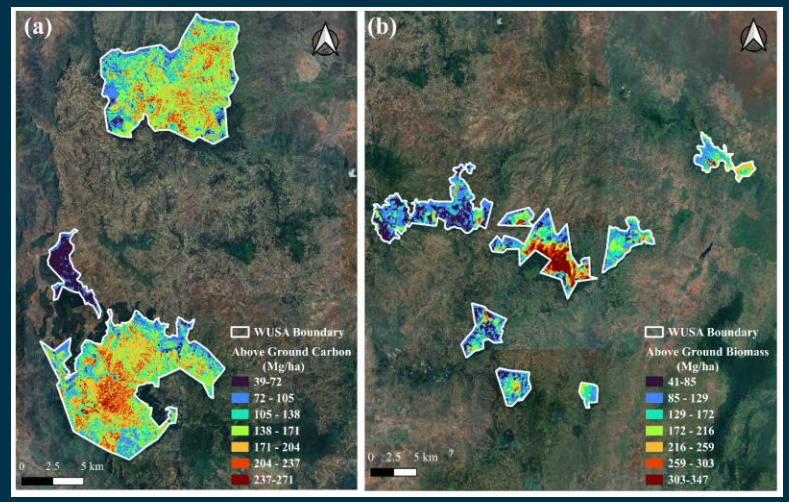
- Higher canopy height observed in dense forest cores

- Lower values in degraded or fragmented areas

(AGB)



(AGC)

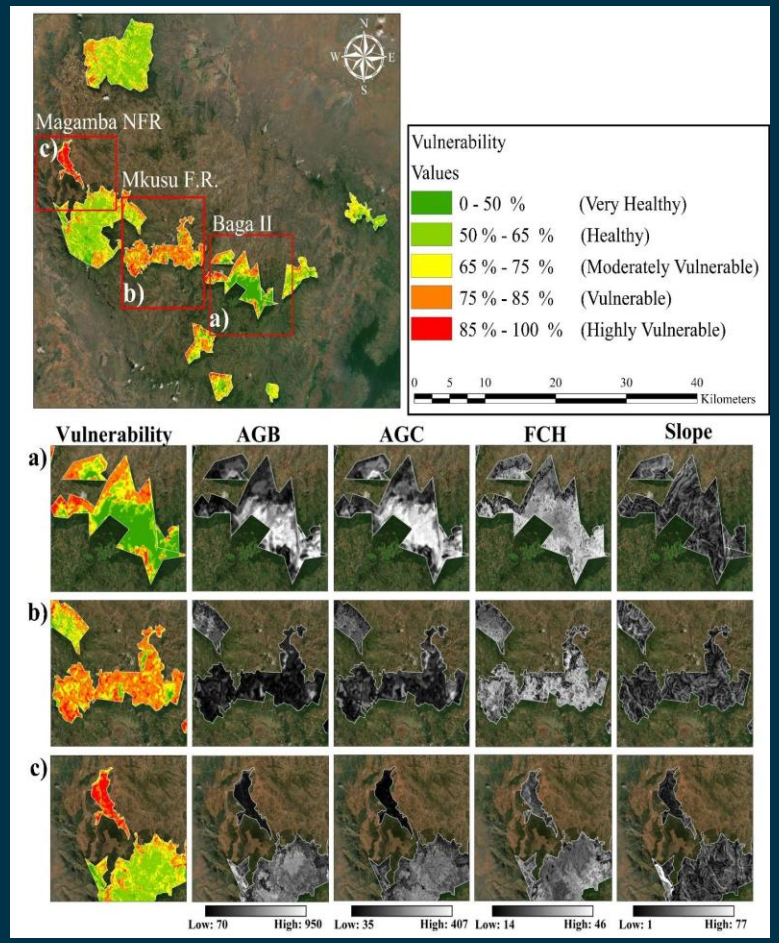


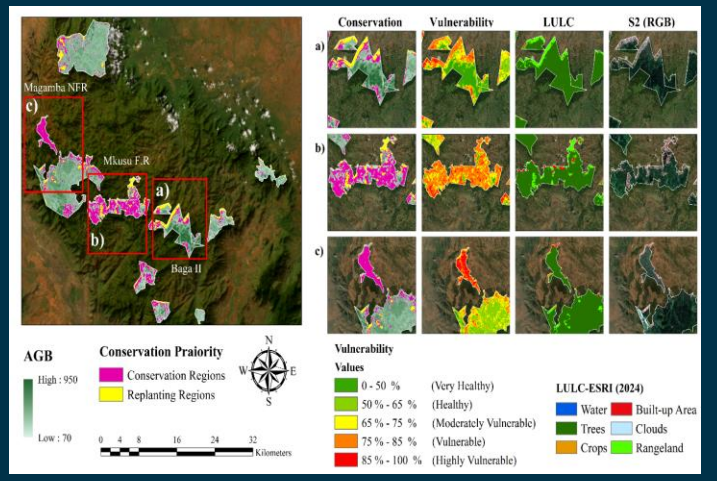
Results

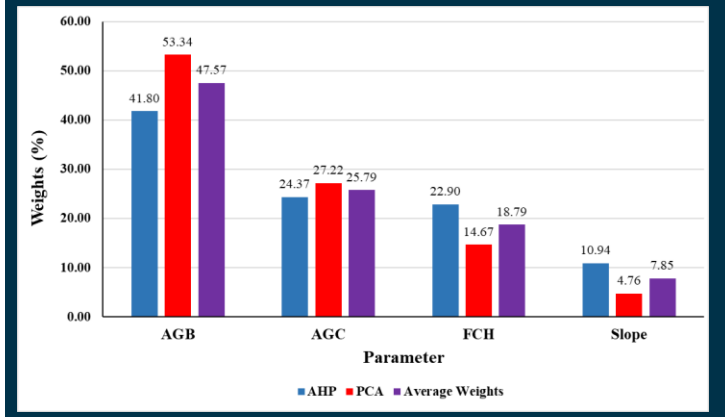
Vulnerability & Degradation Risk Mapping

- Developed **forest vulnerability map** based on:
  - AGB (47.6%)
  - AGC (25.8%)
  - FCH (18.8%)
  - Slope (7.9%)

- Identified:
  - High-risk degradation zones
  - Areas with strong restoration potential







Results

Uncertainty Analysis

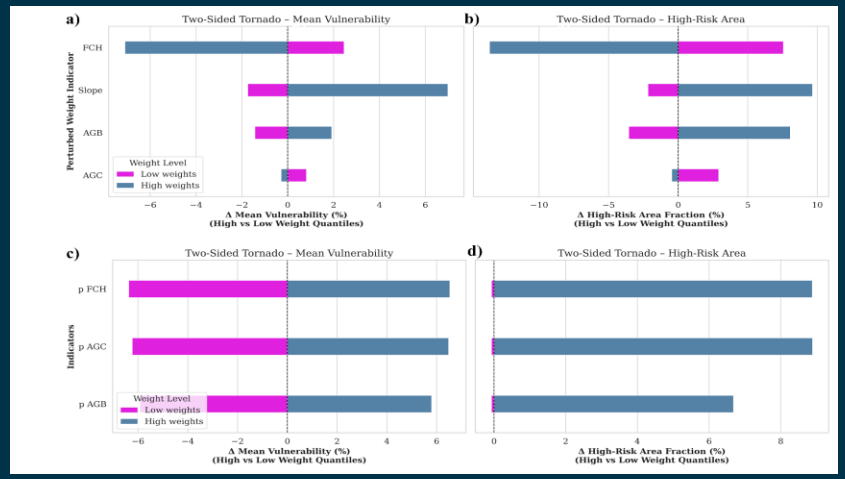
•Applied:

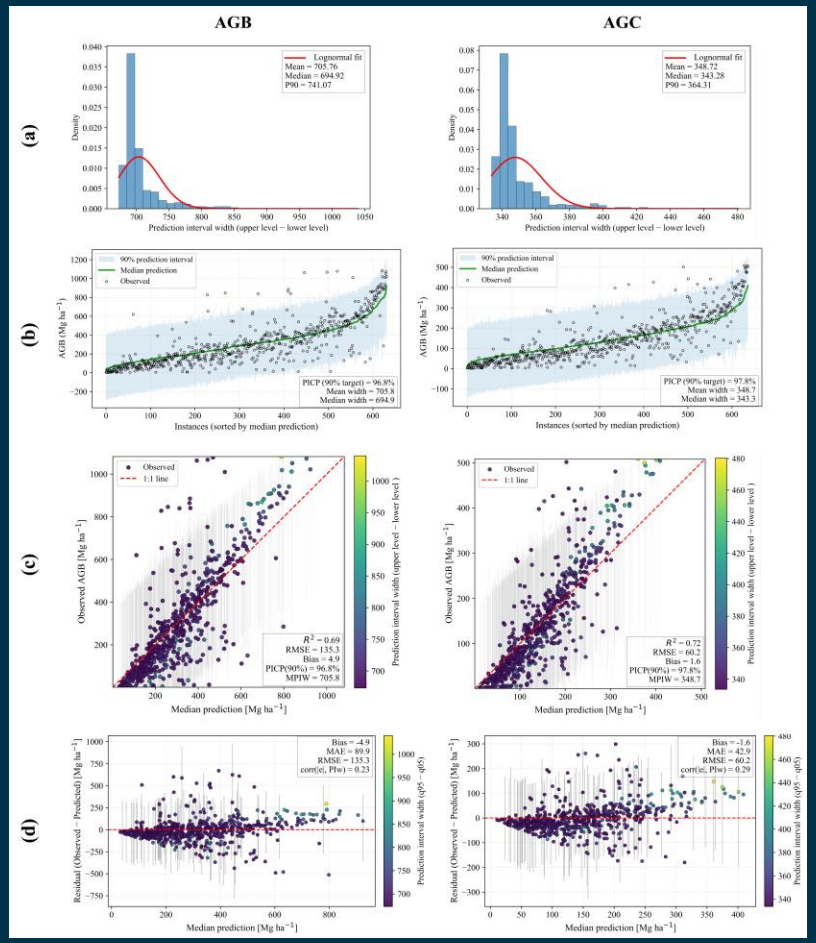
- **Conformal prediction (MAPIE)**
- **Monte Carlo simulation**

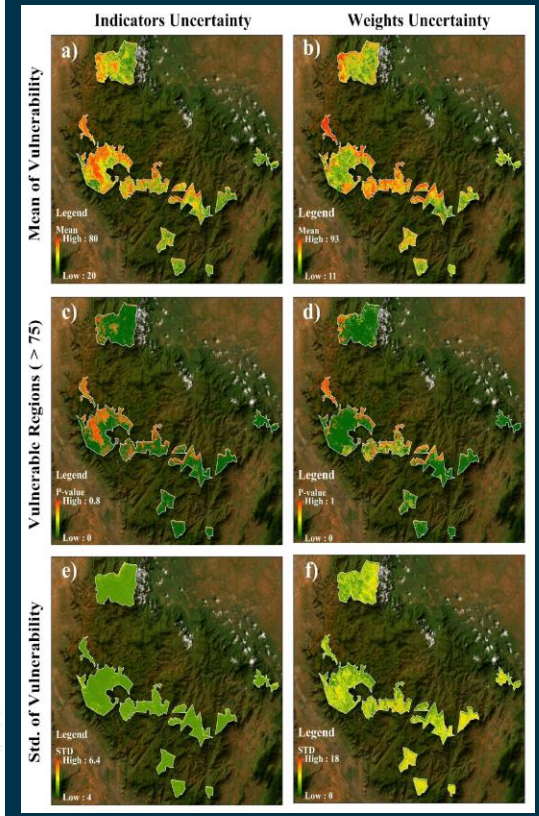
•Findings:

- Model uncertainty strongly influenced by:
  - Canopy height
  - Topography

- Carbon estimates showed relatively lower sensitivity

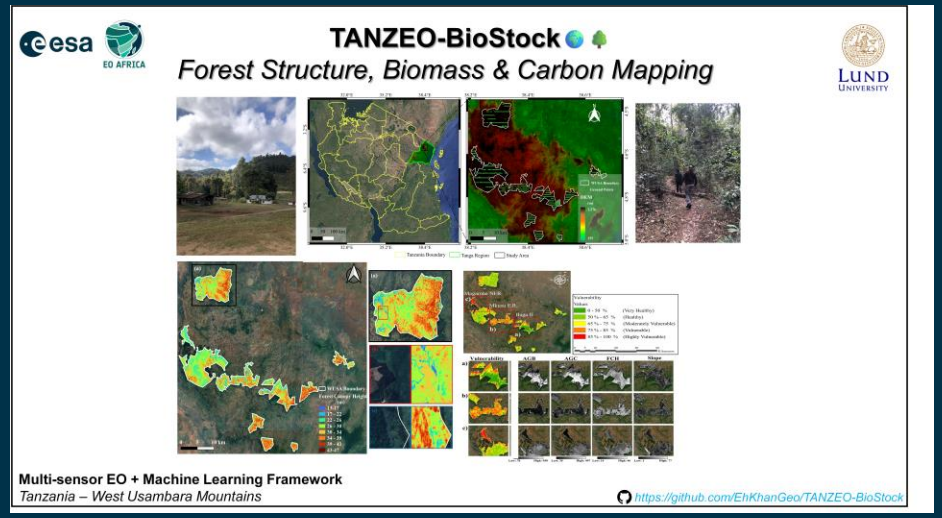






Layout for social media posts

Recommended image resolution 1200 x 675 pixels - landscape format



Post for BS and X (max 230 characters including hashtags)

🌍🌳 TANZEO-BioStock maps forest canopy height, biomass & carbon in Tanzania using ICESat-2, GEDI & ML
Open & reproducible EO framework for climate & forest monitoring
🔗 github.com/EhKhanGeo/TANZEO-BioStock
#RemoteSensing #GIS #Climate

Post in LinkedIn: 600-630 characters including links

🌍🌳 TANZEO-BioStock advances forest monitoring in Tanzania by integrating ICESat-2, GEDI, Sentinel data, and machine learning to map forest canopy height (FCH), biomass (AGB), and carbon stocks (AGC) at 10 m resolution. The framework enables wall-to-wall mapping and supports climate monitoring, sustainable forest management, and food security. Funded by EO AFRICA (ESA & AUC), the project promotes open and reproducible research.
🔗 https://github.com/EhKhanGeo/TANZEO-BioStock