Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institut für Geodäsie und Photogrammetrie

Chunyang Gao & Dominik Senti

ICESat-2 spaceborne LiDAR as complementary data source for biomass mapping

Semester Project

Institute of Geodesy and Photogrammetry
Swiss Federal Institute of Technology (ETH) Zurich

Supervision

Ghjulia Sialelli
Arno Rüegg
Prof. Konrad Schindler

June 2024

Contents

Abstract **1**

1 Introduction **2**

2 Data **4**

2.1 Datasets overview 4

2.2 GEDI data products 4

2.3 Sentinel-2 imagery 5

2.4 ICESat-2 imagery 6

3 Methods **8**

3.1 Research area 8

3.2 Data preprocessing 9

3.3 Data analysis 10

3.4 Data preparation 12

3.4.1 Geographical split 12

3.4.2 Data creation 12

4 Model and Results **15**

4.1 Model 15

4.2 Results 16

5 Conclusion and outlook **17**

5.1 Conclusion 17

5.2 Outlook 17

Abstract

Recent studies emphasize the importance of forest carbon uptake in the global carbon sink, noting significant uncertainties about its scale and distribution. Improving forest monitoring is essential for understanding the carbon cycle and conserving these crucial sinks to help mitigate global warming. The PRS group is developing a satellite-based global biomass monitoring system. Sialelli (2023) evaluates the performance of ML methods to estimate AGBD from optical satellite imagery, leveraging Multi-task learning (MTL), Bayesian learning, and deep ensembles. In line with previous findings, it emphasizes that satellite imagery can be effectively used to estimate Above Ground Biomass Density (AGBD) on a global scale. As part of this effort, the project investigates how to best integrate satellite-based laser altimetry data from the ICESat-2 mission into the biomass mapping system. The current, deep learning-based system fuses GEDI LiDAR data acquired from the International Space Station (ISS) with high-resolution satellite images Sentinel-2 from the ESA Copernicus program, and makes it possible to densely map biomass world-wide. A limitation in this context is the lack of reference data in the sub-arctic region, which lies outside of the ISS orbit. The ICESat-2 mission, originally launched to measure elevation changes in the Earth's cryosphere, captures LiDAR profiles over the polar region (as well as the temperate and tropical regions), and therefore offers the possibility to close this gap. The study finds ways to integrate ICESat-2 observations into the biomass mapping system.

Chapter 1

Introduction

Since the end of the last ice age, 10,000 years ago, the world has lost one-third of its forests (Kump et al. (2004)). Figure 1.1 shows the spatial distribution of deforestation. Tropical forests, which house more than half of the world's species and serve as a significant carbon sink, are profoundly impacted, resulting in the loss of biodiversity and intensifying the climate crisis.

Recent studies have highlighted the significant role that forest carbon uptake plays in enhancing the global carbon sink. Despite its critical importance, there remains substantial uncertainty regarding the scale and distribution of carbon uptake by forests (Pan et al. (2011); Pugh et al. (2020)). Enhancing forest monitoring practices is crucial not only for deepening our understanding of the global carbon cycle but also for implementing strategies to conserve these vital carbon sinks, thereby helping mitigate global warming (Pugh et al. (2019); Liu et al. (2021)). Above-ground biomass estimation, and especially forest biomass, has received considerable attention over the last few decades (Kumar and Mutanga (2017)).

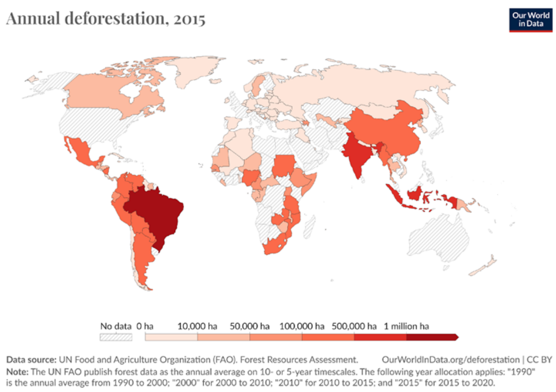

Figure 1.1: Annual Deforestation, 2015.

The standard method for forest carbon stock assessment involves manually recording sample tree metrics, such as diameter at breast height, height, and species at the project site. These measurements are used in forest-specific regression models, known as allometric equations, to estimate the Above Ground Biomass (AGB) and carbon stocks. These models are developed using calibration data obtained through a destructive process where trees are harvested, dried, and weighed. The other approach is to leverage remote sensing (RS) technologies to estimate the tree parameters used as input for the allometric equations. Instead of relying on allometric equations, machine learning (ML) methods,on the other hand, can utilize globally accessible satellite imagery to reduce the time and cost of producing biomass maps and improve accuracy. It has been used widely in biomass estimation, (Sialelli (2023))

Some efforts have been made to map AGB using RS-based ML methods:

- Lang et al. (2022) introduces a new supervised machine learning strategy to analyze GEDI waveforms and estimate global canopy top height, utilizing a probabilistic deep learning framework with a group of deep convolutional neural networks (CNNs) to circumvent the direct modeling of indeterminate factors like atmospheric noise. The model is capable of identifying strong features that are applicable across different unseen geographic areas and provides dependable predictions of uncertainty. It consistently estimates global canopy top height with an expected root mean square error (RMSE) of 2.7 meters and maintains low bias.
- Sialelli (2023) estimates above ground biomass using optical satellite imagery and LiDAR-based reference data. Building on prior studies, it introduces the innovation of employing Multi-Task Learning (MTL) to learn exclusively from satellite data. The study enhances the interpretability of the predictions by quantifying estimation uncertainties through the use of Bayesian deep ensembles.

Our work is built on Sialelli (2023). It can densely map biomass world-wide. A limitation is the lack of reference data in the sub-arctic region, which lies outside of the ISS orbit. The project aims to integrate ICESat-2 observations into the biomass mapping system and assess its contribution to the mapping accuracy.

The structure for the report is as follows. In Chapter 2, the three different datasets used in the project are introduced. In Chapter 3, the method and research area for the project are presented. The data preprocessing steps, data analysis visualization and machine learning architecture used are introduced. In Chapter 4, the results are shown and analyzed. Chapter 5 provides the conclusion and outlook of the project.

Chapter 2

Data

2.1 Datasets overview

We use three different datasets in our project: GEDI Data Products, Sentinel-2 Imagery and ICESat-2 data. Table 2.1 shows their main characteristics: spatial coverage, file format and spatial resolution.

Characteristics	GEDI	Sentinel-2	ICESat-2
Spatial coverage	Lat: 51.6°N–51.6°S	Global	Lon: -179°–178°, Lat: 43°–78°
File format	Point shapefile	GeoTIFF	GeoTIFF
Spatial resolution	60 m footprints	10 m, 20 m, and 60 m	30 m grid cells

Table 2.1: Comparison of characteristics of GEDI, Sentinel-2, and ICESat-2 data products

2.2 GEDI data products

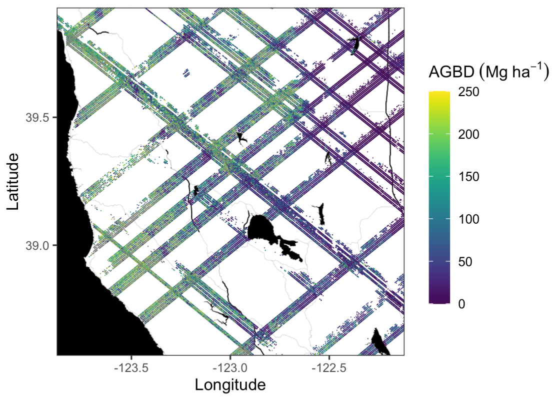

Figure 2.1: Example subset of aboveground biomass density predictions from the GEDI Level-4A footprint product over Northern California, U.S., spanning April to July 2019 (Dubayah et al. (2022))

The Global Ecosystem Dynamics Investigation (GEDI) (see Figure 2.1)produces high resolution laser ranging observations of the 3D structure of the Earth. GEDI's precise measurements of forest canopy height, canopy vertical structure, and surface elevation greatly advance our ability to characterize important carbon and water cycling processes, biodiversity, and habitat. The footprints are located within the global latitude band observed by the International Space Station (ISS), nominally 51.6 degrees N and S and reported for the period 2019-04-18 to 2020-09-02. The GEDI instrument consists of three lasers producing a total of eight beam ground transects, which instantaneously sample eight ∼25 m footprints spaced approximately every 60 m along-track. The GEDI beam transects are spaced approximately 600 m apart on the Earth's surface in the cross-track direction, for an across-track width of ∼4.2 km. Footprint AGBD was derived from parametric models that relate simulated GEDI Level 2A (L2A) waveform relative height (RH) metrics to field plot estimates of AGBD. Height metrics from simulated waveforms associated with field estimates of AGBD from multiple regions and plant functional types (PFT) were compiled to generate a calibration dataset for models representing the combinations of world regions and PFTs (i.e., deciduous broadleaf trees, evergreen broadleaf trees, evergreen needleleaf trees, deciduous needleleaf trees, and the combination of grasslands, shrubs, and woodlands). (Dubayah et al. (2022))

2.3 Sentinel-2 imagery

Sentinel-2 is a European mission focused on high-resolution, multi-spectral imaging, with a revisit frequency of five days at the Equator. The resulting images are 100×100 km^2 orthoimages in UTM/WGS84 projection, as shown in Figure 2.2. Sentinel-2 offers two main products: Level-1C (which provides top-of-atmosphere reflectances in cartographic geometry) and Level-2A (which offers bottom-of-atmosphere reflectances in cartographic geometry). The latter includes 12 spectral bands ranging from visible to shortwave infrared, with spatial resolutions reaching up to 10 meters per pixel, detailed in Table 2.2. Additional outputs include an Aerosol Optical Thickness (AOT) map, a Water Vapour (WV) map, and a Scene Classification (SCL) map, along with Quality Indicators (QI) for cloud and snow probabilities at a 60-meter resolution. Due to its frequent revisit times, Sentinel-2 data is crucial for monitoring environmental changes on a global scale and has become an essential tool for the remote sensing community.

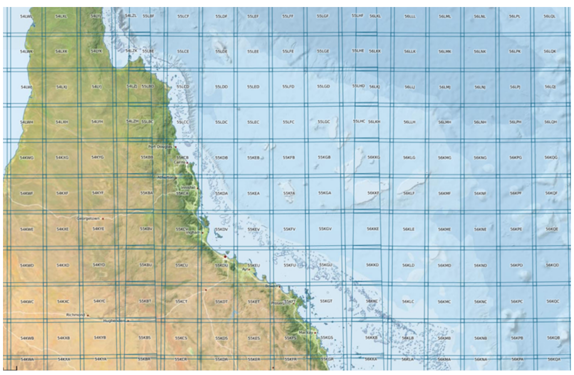

Figure 2.2: Sample Sentinel-2 UTM Tiling Grid, ESA (eAtlas (2024))

Band	Resolution	Description
B1	60 m	Aerosols
B2	10 m	Blue
B3	10 m	Green
B4	10 m	Red
B5	20 m	Red Edge 1
B6	20 m	Red Edge 2
B7	20 m	Red Edge 3
B8	10 m	Near Infrared (NIR)
B8a	20 m	Red Edge 4
B9	60 m	Water vapor
B11	20 m	Short Wave Infrared (SWIR) 1
B12	20 m	SWIR 2

Table 2.2: Sentinel-2 Spectral Bands and Resolutions

2.4 ICESat-2 imagery

The ICESat-2 (Ice, Cloud, and land Elevation Satellite-2) mission, equipped with the Advanced Topographic Laser Altimeter System (ATLAS), measures Earth's surface heights using 10,000 laser pulses per second. This provides detailed elevation data for sea ice, land ice, forest canopies, water bodies, urban areas, and more, with data collection starting from late 2018. Combined with data from previous missions like ICESat/GLAS and Operation IceBridge, ICESat-2 helps researchers monitor cryosphere changes, assess sea ice thickness, track iceberg movements, and predict the impact of melting glaciers on global sea level rise.

Additionally, ICESat-2's ATLAS measures heights of ocean and land surfaces, including forests, snow, lakes, rivers, ocean waves, and urban areas. This data is valuable for various scientific disciplines and hazard applications, such as estimating global forest carbon storage, improving wildfire behavior forecasts, and predicting volcanic eruptions.

The ICESat-2 Aboveground Dry Woody Biomass Density (AGBD) dataset (Duncanson et al. (2023)) (see Figure 2.3), provided by the NSIDC DAAC, offers 30-meter resolution estimates for high northern latitude forests. It is designed both for boreal-wide mapping and filling the northern spatial data gap from NASA's Global Ecosystem Dynamics Investigation (GEDI) project. This dataset utilizes Ordinary Least Squares (OLS) regression and random forest models, combined with Harmonized Landsat Sentinel-2 (HLS) and Copernicus DEM data, to predict AGBD across 90-km tiles. The dataset includes cloud-optimized GeoTIFF files with mean AGBD and standard deviation estimates.

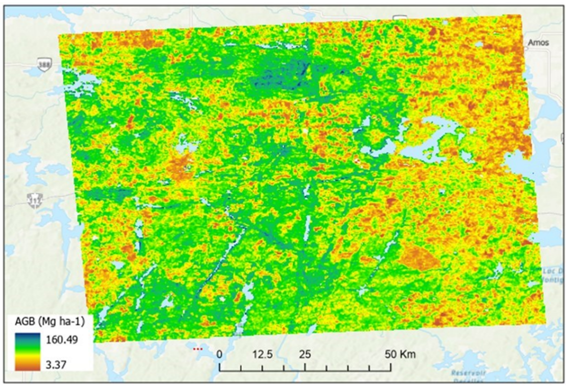

Figure 2.3: Aboveground biomass (Mg ha^{-1}) for boreal forest estimated from ICESat-2 imagery over the city of Rouyn-Noranda in western Ontario, Canada (approximately -78.758 longitude, 48.259 latitude)

Chapter 3

Methods

3.1 Research area

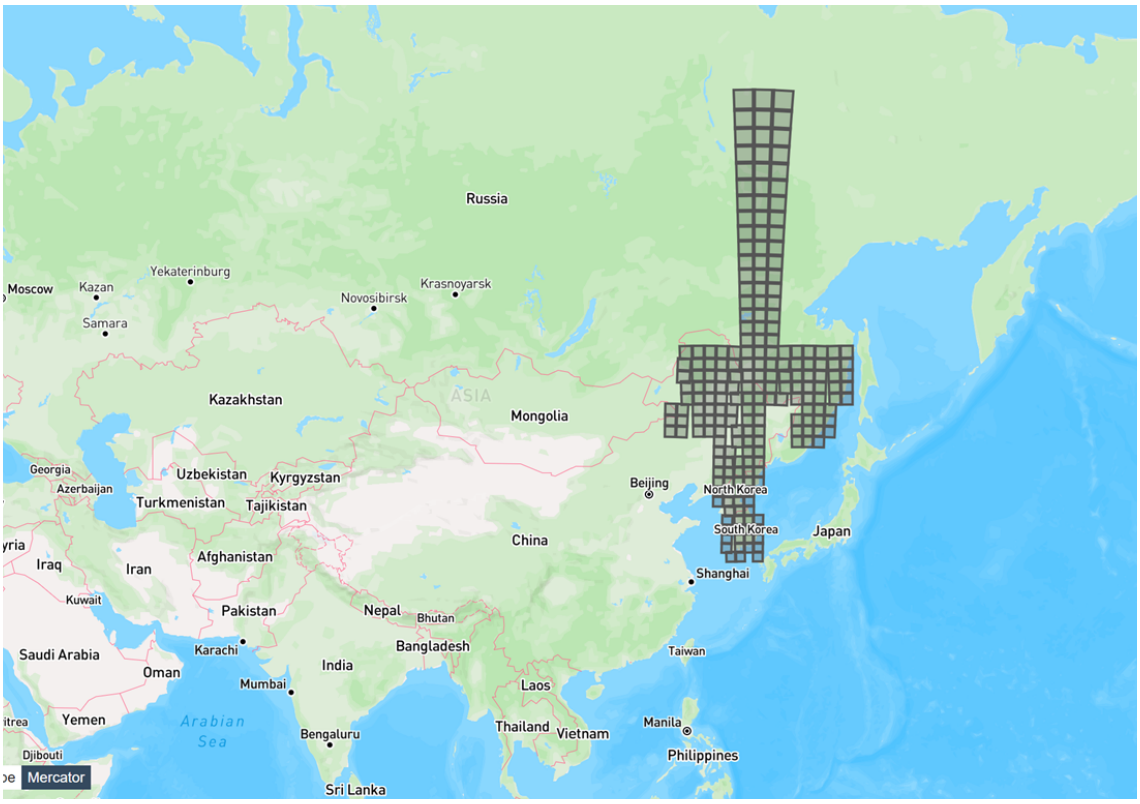

Figure 3.1: Research area defined by bounds of Sentinel-2 tiles

The Figure 3.1 shows the bounds of Sentinel-2 tiles, covering the geographic region including Siberia. These tile bounds serve as the basic units for the study, defining the spatial scope. The spatial scope includes sub-arctic regions that have corresponding ICESat-2 AGBD datasets, which serves as the base to integrate ICESat-2 observations into the biomass mapping system.

The Figure 3.2 illustrates the coverage of three different datasets. Red dots represent the coverage

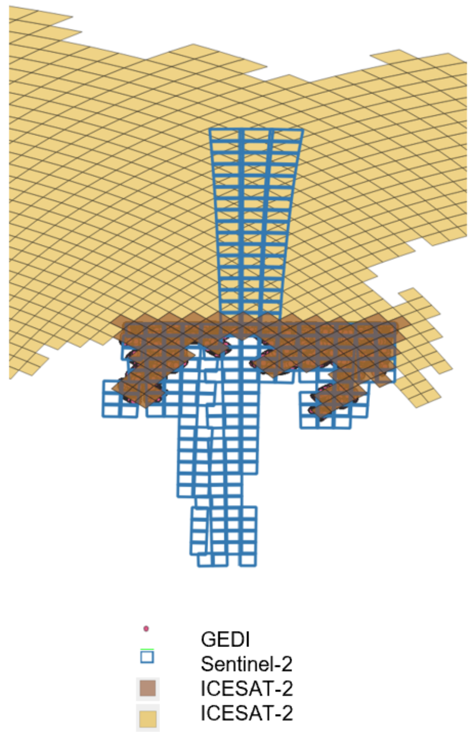

Figure 3.2: Spatial coverage of different datasets

of the GEDI dataset. Brown and yellow squares represent coverage areas of the ICESat-2 dataset. Brown squares also represent the regions where these three datasets overlap. Different overlapping areas are used for training and inference separately.

3.2 Data preprocessing

The Figure 3.3 illustrates the ICESat-2 data processing workflow. Initially, the ICESat-2 data that overlap with the given Sentinel-2 data are identified using shapefile and the corresponding names of the ICESat-2 GeoTIFF files are saved. Subsequently, these ICESat-2 GeoTIFF files are downloaded using the download links based on these overlapping images. The downloaded data that intersect with the same Sentinel-2 tile are merged into a single large mosaic image. This merged mosaic is then reprojected to the coordinate system of their corresponding Sentinel-2 tile. Finally, the reprojected mosaic images are cropped to the bounds of the Sentinel-2 tiles. After data preprocessing, the ICESat-2 datasets can be used for creating patches and subsequently training the model.

We do not consider AGBD values that are bigger than 500 both in Sentinel-2 and ICESat-2, as the literature by Carreiras et al. (2017) suggests, to maintain a representative sample of values (Sialelli (2023)).

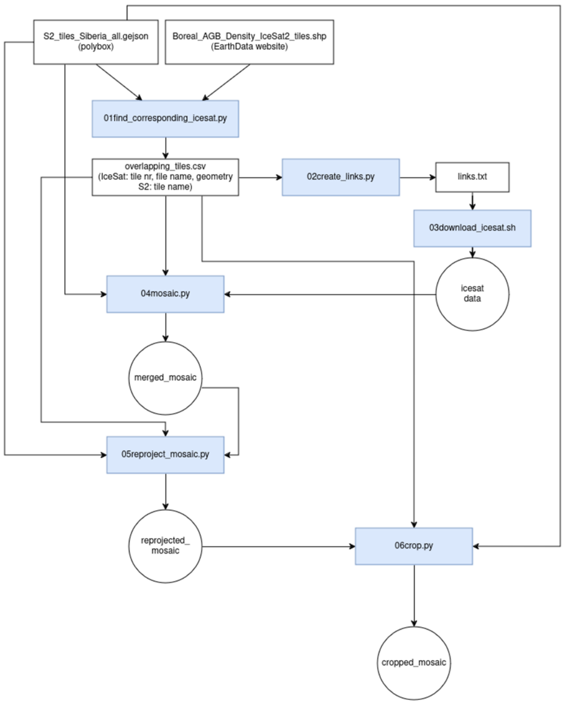

Figure 3.3: Flow chart for preprocessing ICESat-2 dataset

3.3 Data analysis

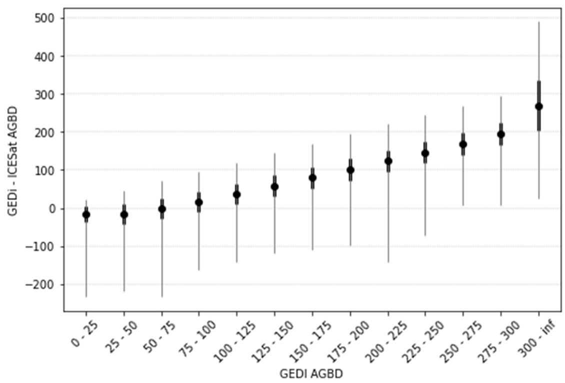

Figure 3.4: The boxplot of GEDI-ICESat AGBD (on the y axis) for every bin of the GEDI AGBD (x axis).

Before training the model, we compare the the biomass density values of GEDI products and ICESat-2 products to have an overview of our datasets.

Figure 3.4 reveals the distribution and variability of discrepancies in GEDI and ICESat-2 Above Ground Biomass Density (AGBD) across different GEDI biomass density categories. The y-axis measures the differences in AGBD between GEDI and ICESat-2 data across GEDI biomass density bins shown on the x-axis. The plot shows an increasing trend in the median values of GEDI-ICESat AGBD as the GEDI AGBD increases. This trend is especially noticeable in the higher AGBD bins (200 and above). In the lower AGBD bins (e.g., 0-25), the variability of GEDI-ICESat AGBD is smaller, as indicated by shorter error bars. In contrast, the higher AGBD bins (e.g., 275-300, 300-inf) exhibit greater variability, indicated by longer error bars. This suggests that in higher AGBD bins, the GEDI-ICESat AGBD measurements are more dispersed, which means that ICESat-2 tends to underestimate AGBD for higher GEDI AGBD.

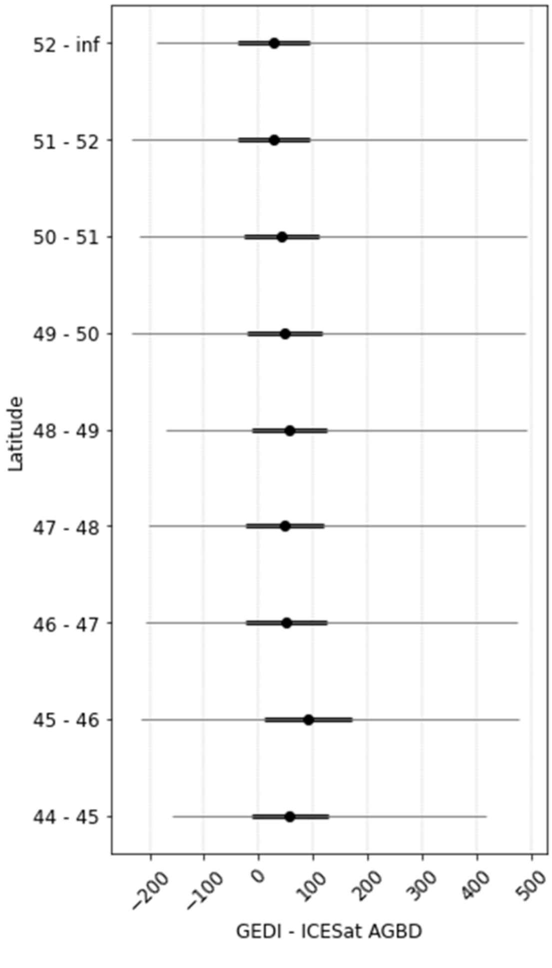

Figure 3.5: Box plot of GEDI-ICESat AGBD (on the x axis) for every latitude bin (y axis).

Figure 3.5 reveals the distribution and variability of discrepancies in GEDI and ICESat-2 Above Ground Biomass Density (AGBD) across different latitude bins. From the plot, we can see that the medians across different categories appear to be very close to zero, suggesting that the AGBD differences between GEDI products and ICESat-2 are small. The variability (spread) of AGBD differences within each category seems to be consistent, with whiskers of approximately similar length for all categories. It also can be concluded that ICESat-2 tends to underestimate AGBD compared to GEDI. The difference distribution has little correlation with latitudes.

More figures for GEDI and ICESat-2 AGBD analysis can be found in Appendix.

3.4 Data preparation

3.4.1 Geographical split

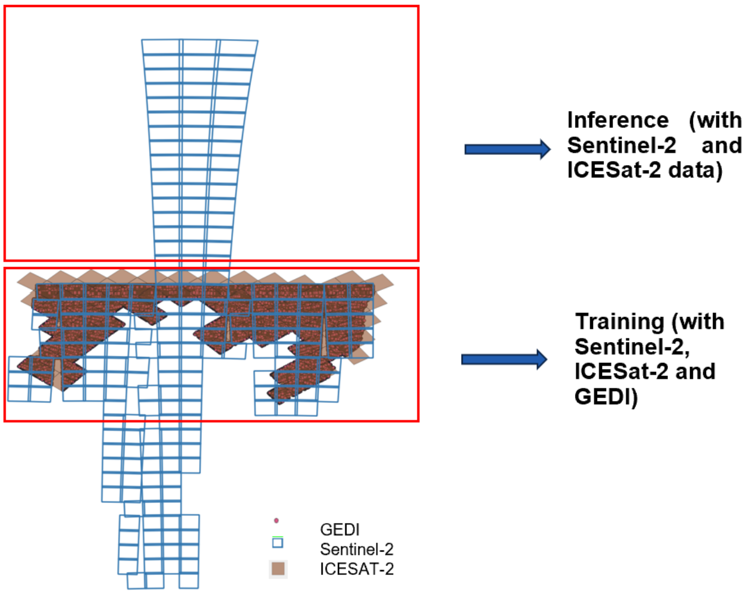

Figure 3.6: Dataset scope used for training and inference

Figure 3.6 illustrates the division of regions for training and inference based on different overlapping areas. The regions with Sentinel-2, ICESat-2 and GEDI datasets are used for training while the regions with only Sentinel-2 and ICESat-2 datasets are used for inference.

The datasets used for training in Figure 3.6 are shuffled and then separated into train/validation/test sets. All GEDI footprints that belong to the same tile, belong to the same set (train/validation/test). The ICESat-2 GeoTIFF files are mosaicked and cropped to the Sentinel-2 bounds. Figure 3.7 shows the regions for train, validation and test.

3.4.2 Data creation

We utilize a patch-based training method, with each patch sized at 15×15 pixels, corresponding to $150 \times 150\text{m}^2$ on the ground. Each patch contains one valid ground truth pixel at its center. Pixels lacking ground truth do not contribute to the per-patch loss, ensuring they do not affect the training process. Here is a brief introduction for the patches creating steps.

- Command-line arguments are parsed to obtain necessary information such as patch size, and input/output paths.
- Data Loading and Filtering: Sentinel-2 tile names are listed. The geometries of these tiles are read from a Sentinel-2 grid shapefile.

  GEDI footprints intersecting the geometry of Sentinel-2 tiles are loaded. Unnecessary columns are removed, and the remaining data is reprojected to match the CRS of the Sentinel-2 tiles.

  ICESat-2 data for the specified year and tile are loaded and prepared for patch extraction.

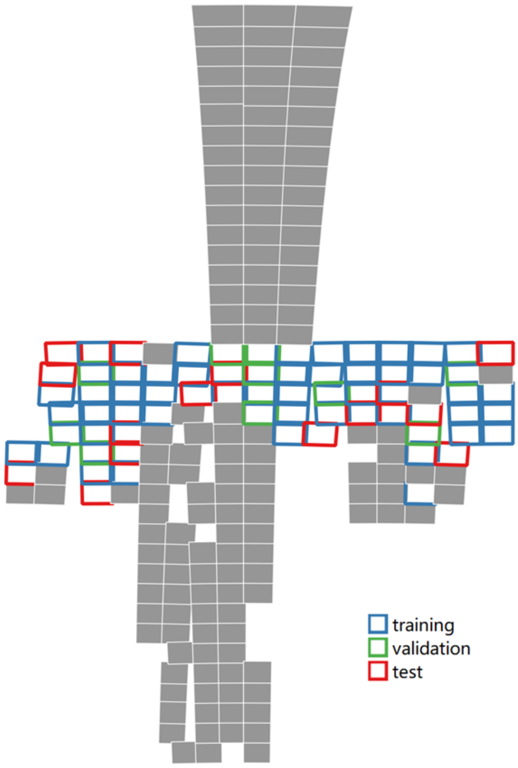

Figure 3.7: Geographical split of datasets used for training

- Reprojection and Cropping: GEDI data is reprojected to the same CRS as Sentinel-2 data.

  Sentinel-2 bands are reprojected, upsampled to a 10m resolution, and corrected for radiometric offsets if needed.

- Grouping and Matching: GEDI footprints are grouped by acquisition date.

  For each group, the corresponding Sentinel-2 L2A product closest in time to the GEDI footprint is identified.

- Patch Extraction: For each GEDI footprint, image patches centered on the footprint are extracted from Sentinel-2 and ICESat-2 data. The quality of patches is checked to ensure the center pixel is vegetated.

  Sentinel-2 data patches are created for various spectral bands and scene classification layers, while ICESat-2 data patches are extracted and resampled to align with Sentinel-2 data.

- Data Transformation and Upsampling: Sentinel-2 and ICESat-2 data patches are transformed to maintain consistent resolution and format. Missing values in the patches are filled using nearest-neighbor interpolation.

  Attributes like acquisition date, processing baseline number (PBN), and relative orbit number (RON) are extracted and added to the data.

- Data Aggregation: The extracted patches and associated metadata from Sentinel-2, GEDI, and ICESat-2 datasets are aggregated into structured dictionaries.

  Aggregated results are stored in HDF5 files, structured to accommodate the different datasets and attributes.

- Parallel Processing: The entire processing task is divided into multiple parts, with each part handled by a separate process in parallel to improve efficiency.

Each parallel process handles its assigned Sentinel-2 tiles sequentially, extracting patches and writing results to corresponding output files.

- Output and Cleanup: Processed data is saved to HDF5 files in the specified output directory. The structure includes groups for each tile and datasets for Sentinel-2 bands, GEDI attributes, and ICESat-2 data.

  Temporary files, such as unzipped Sentinel-2 products, are deleted after processing to free up storage space.

Chapter 4

Model and Results

4.1 Model

The model architecture that is experimented with is a Fully Convolutional Network (FCN) as shown in Figure 4.1. It is a particular type of CNNs that takes inputs of arbitrary size and produces outputs of the same size. The FCN used in the project consists of seven stacked convolutional blocks (see Figure 4.2).

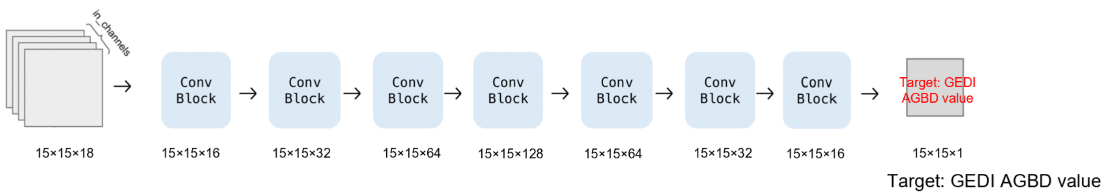

Figure 4.1: Architecture of the Fully Convolutional Neural network (FCN)

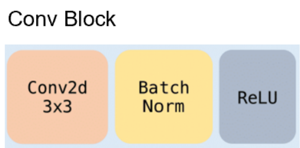

Figure 4.2: Convolutional block

Feature sources	Features
Sentinel-2	12 spectral bands
Position	Cos/sin latitude/longitude
ICESat-2	AGB, standard deviation

Table 4.1: Feature sources and their corresponding features

Patches of 15 × 15 pixels (corresponding to 150 × 150 m^2 on the ground) both from the Sentinel-2 mosaic and the ICESat-2, centered around GEDI footprints are used as input for the machine learning task. Each patch has eighteen features as shown in Table 4.1. The input features include 12 spectral bands from Sentinel-2 dataset, position information from latitude and longitude and AGB band, standard deviation band from ICESat-2 dataset. The output is the patch of 15 $\times$ 15 pixels with estimated AGB values in the central pixel. The GEDI AGBD values serve as the reference AGBD.

The optimisation of the parameters of the networks is done using ADAM. It can adaptively adjusts the learning rate for each trainable parameter by normalizing the global learning rate with the running average of the gradient. The Root Mean Squared Error (RMSE) is used as loss function. The base learning rate is set to 0.001, and the batch size to 512 patches. Only the predicted AGBD of the central pixel is backpropagated to optimize the model parameters.

We utilize different random seeds to obtain varied initialization parameters, resulting in five distinct models. By averaging the predictions from these models, we derived the final biomass prediction. This ensemble approach helps to mitigate the variability and uncertainty inherent in individual model predictions, leading to more robust and reliable biomass density estimates.

For inference, only the AGBD values of regions with ICESat-2 and Sentinel-2 data are predicted using the trained model. There are no reference GEDI AGBD values for evaluating the model performance. The average of Sentinel-2 tiles in different timestamps are used as Sentinel-2 data input.

4.2 Results

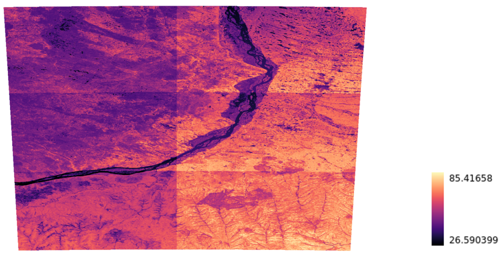

Figure 4.3: Example biomass inference result in the sub-arctic region

Due to time and resource limit, we can only visualize models' predictions on a couple of selected Sentinel-2 tiles as shown in Figure 4.3. The map currently shows some artefacts at the borders of the Sentinel-2 tiles. They could be smoothed to improve the visual appearance of the map.

Chapter 5

Conclusion and outlook

5.1 Conclusion

In this project, we investigated the integration of ICESat-2 spaceborne LiDAR data into the satellite-based global biomass monitoring system developed by Sialelli (2023). The primary goal is to enhance the accuracy and spatial coverage of Above Ground Biomass Density (AGBD) estimation, especially in sub-arctic regions that lack sufficient reference data.

Our approach involved preprocessing and analyzing data from three different sources: GEDI, Sentinel-2, and ICESat-2. The Fully Convolutional Network (FCN) model is employed to integrate these diverse data sources to predict biomass density. GEDI AGBD values served as ground truth. Sentinel-2 and ICESat-2 bands are used as input features to predict biomass density.

A significant portion of the project is dedicated to the data preprocessing phase and preparing the data for machine learning. Considerable time and effort are spent attempting to harmonize data from these three disparate sources. Consequently, we have conducted fewer experiments with different model architectures, and the inference phase does not encompass a broader range of study area. Despite these limitations, the results demonstrated that the inclusion of ICESat-2 data can help provide reference data in sub-arctic region. This integration allowed for better mapping of biomass in areas previously underserved by other satellite missions.

In conclusion, our study highlights the potential of combining ICESat-2 data with other satellite-based observations to enhance global biomass monitoring efforts. The methodology and findings presented in this project contribute valuable insights towards developing more accurate and comprehensive biomass maps, which are crucial for understanding and mitigating the impacts of global warming.

5.2 Outlook

- Exclude biomass outliers: Rüegg (2022) showed that high canopies are measured with strong uncertainty constraints. Since these very high observations are assumed to be outliers, Arno defined a threshold-function for the maximal canopy height based on the latitude. With regard to our study, we can compute the corresponding biomass density threshold using the formula to compute the biomass from the canopy height to exclude outliers and acquire more reliable results.

- Combine more data sources: Sialelli (2023) also uses canopy height as input for estimating biomass. The canopy height data provides information about the vertical structure of the forest, while the Sentinel-2 data provides information about the spectral characteristics of the forest. By combining these two sources of information, the models are able to better capture the complexity of the relationship between these variables, resulting in improved performance. Given this work, it is also meaningful to include canopy height data to our input features.

- Experiment complex model architecture: In this project, we used only the Fully Convolutional Network (FCN) for model training and inference. However, future research could explore incorporating other more complex network architectures, such as ResNeXt and Xception. These advanced neural network models may further improve the accuracy and generalization ability of biomass density estimation, enhancing our capacity to understand and monitor forest biomass across different geographic regions.

- Remove artefacts: The current results show artefacts, which can disrupt the biomass maps. To address this, we can apply methods to create a smooth transition between neighboring tiles. Techniques such as blending, interpolation, and edge-smoothing algorithms can be used to mitigate these artefacts.

- Work with big data: Our research area currently includes only a portion of the sub-arctic region. To obtain more comprehensive and accurate global biomass estimations, it is essential to integrate data from a broader range of geographical areas.

- Comparison with other biomass density maps: When doing inference in high-latitude regions, we predict biomass density from Sentinel-2 and ICESat-2 datasets without reference data. By comparing these predictions with existing biomass density maps, we can identify discrepancies, validate our models, and refine our methodologies to ensure more accurate and dependable biomass estimations.

Bibliography

Carreiras, J. M., Quegan, S., Le Toan, T., Minh, D. H. T., Saatchi, S. S., Carvalhais, N., Reichstein, M., and Scipal, K. (2017). Coverage of high biomass forests by the esa biomass mission under defense restrictions. *Remote Sensing of Environment*, 196:154–162.

Dubayah, R., Armston, J., Kellner, J., Duncanson, L., Healey, S., Patterson, P., Hancock, S., Tang, H., Bruening, J., Hofton, M., Blair, J., and Luthcke, S. (2022). Gedi l4a footprint level aboveground biomass density, version 2.1. Accessed: 2022-06-13.

Duncanson, L., Montesano, P., Neuenschwander, A., Thomas, N., Mandel, A., Minor, D., Guenther, E., Hancock, S., Feng, T., Barciauskas, A., Chang, G., Shah, S., and Satorius, B. (2023). Aboveground biomass density for high latitude forests from icesat-2, 2020.

eAtlas (2024). Sentinel-2 data access and usage. Accessed: 2024-06-13.

Kumar, L. and Mutanga, O. (2017). Remote sensing of above-ground biomass. *Remote Sensing*, 9(9).

Kump, L. R., Kasting, J. F., and Crane, R. G. (2004). *The Earth System*, volume 432. Pearson Prentice Hall, Upper Saddle River, NJ.

Lang, N., Kalischek, N., Armston, J., Schindler, K., Dubayah, R., and Wegner, J. D. (2022). Global canopy height regression and uncertainty estimation from gedi lidar waveforms with deep ensembles. *Remote Sensing of Environment*, 268:112760.

Liu, A., Cheng, X., and Chen, Z. (2021). Performance evaluation of gedi and icesat-2 laser altimeter data for terrain and canopy height retrievals. *Remote Sensing of Environment*, 264:112571.

Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., Phillips, O. L., Shvidenko, A., Lewis, S. L., Canadell, J. G., et al. (2011). A large and persistent carbon sink in the worldâs forests. *science*, 333(6045):988–993.

Pugh, T. A., Lindeskog, M., Smith, B., Poulter, B., Arneth, A., Haverd, V., and Calle, L. (2019). Role of forest regrowth in global carbon sink dynamics. *Proceedings of the National Academy of Sciences*, 116(10):4382–4387.

Pugh, T. A., Rademacher, T., Shafer, S. L., Steinkamp, J., Barichivich, J., Beckage, B., Haverd, V., Harper, A., Heinke, J., Nishina, K., et al. (2020). Understanding the uncertainty in global forest carbon turnover. *Biogeosciences*, 17(15):3961–3989.

Ritchie, H. (2021). Deforestation and forest loss. *Our World in Data*. https://ourworldindata.org/deforestation.

Rüegg, A. (2022). Vegetation mapping with icesat-2.

Sialelli, G. (2023). Global biomass estimation and uncertainty quantification with multi-task bayesian deep ensembles.

Appendix A

GEDI and ICESat-2 AGBD analysis

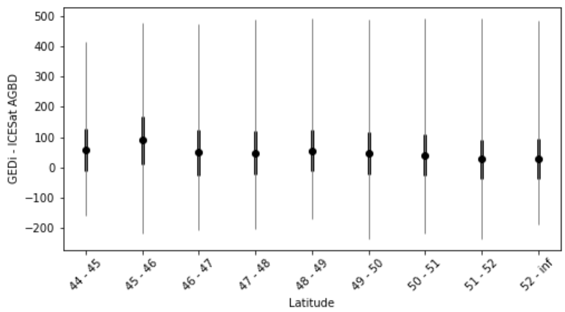

Figure A.1: Box plot of GEDI-ICESat AGBD (on the y axis) for every latitude bin (x axis)

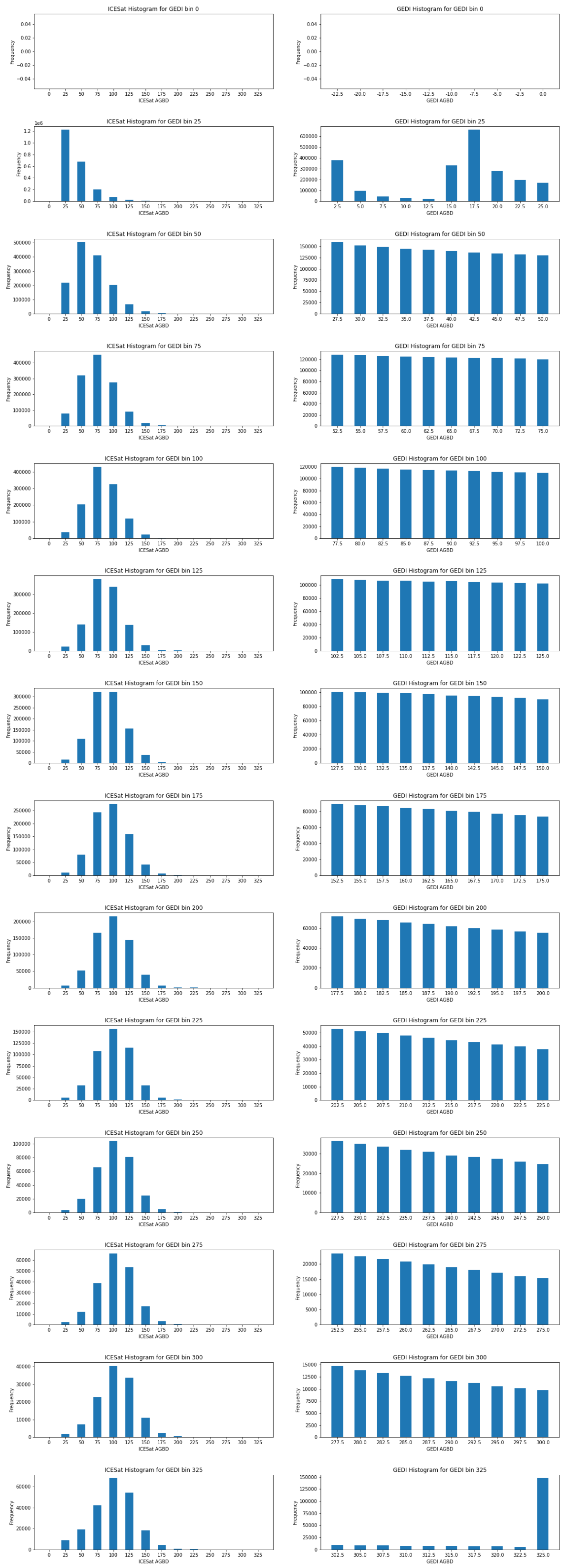

Figure A.2: Comparison of Distribution and Frequency of Biomass Density Estimates Across Different GEDI AGBD bins

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of Originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of Master's thesis (in block letters):

ICESat-2 spaceborne LiDAR as complementary data source for biomass mapping

Authored by (in block letters):
For papers written by groups the names of all authors are required.

Name(s):	**First name(s):**
Gao	Chunyang

With my signature I confirm that
- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	**Signature(s)**
Zurich, 14/06/2024	Chunyang Gao

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.