BioMassters

Estimating the Biomass of the Finnish Forest

The Team

- **Sofia Matias**
  CEO / Data Engineer
- **Pedro Lopes**
  Machine Learning Engineer
- **Emile van Cauwenberghe**
  Machine Learning Engineer
- **Luis Sarmento**
  ML Ops / Data Scientist

Background

Measuring the power of forests as carbon sinks provides incentive to consider the environmental value of forests and contribute to the global goal of halting deforestation.

Biomass:

The total mass of living matter within a given unit of environmental area.

The problem at a glance

Goal: To estimate the yearly biomass of different sections of Finland's forests and its effect on carbon capture.

Idea: Remote sensing methods, such as satellite imagery, offer a much faster, less intrusive, cheaper and more wide spread biomass estimate.

Challenge: Both satellites use a total of 15 different sensors to create a gridded section image of the forest.

Input

- 2 Satellites
- 15 layers (4+11)
- 1 per month, for a year

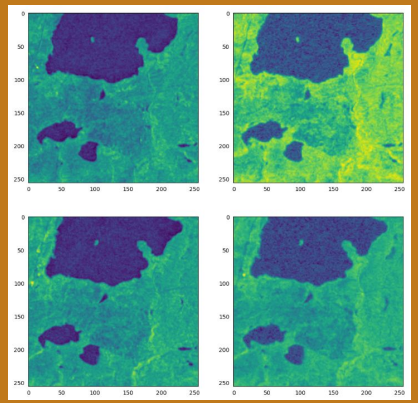
Pictures per patch of land:
(4+11)*12 = 180

- 8689 patches of forest (chips)

A total of: more than **330Gb** of data to process

Sentinel-1. Image credit: ESA Sentinel Online


Sentinel-2. Image credit: ESA Sentinel Online

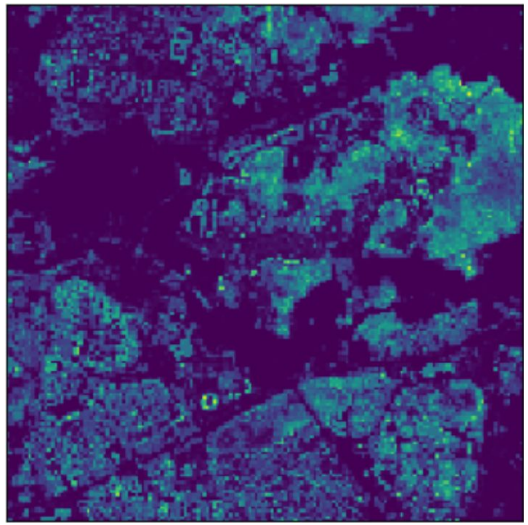


Sentinel-1

Sentinel-2

Output

- 1 LiDAR image per patch of forest
- 8689 LiDAR images
- To be used in the calculation of biomass volume



LiDAR

LiDAR:

Is used to make digital 3-D models of areas on the Earth's surface.

It can be used to calculate the volume of vegetation in the forest - *Biomass*.

Machine Learning

Our first idea was to use Machine Learning to solve the problem.

However, this proved too big of a task due to the **size of the data**.

Limitations

Data size was also a limiting factor in available **processing capacity** of our computers, making it difficult to train the model.

Limitations cont.

Our third challenge came in the form of the **output as an image**. This created a technical issue for our Team.

Finally, there is a **lack of data** for a few months, due limitations like cloud coverage.

The solution

Deep Learning is part of a broader family of machine learning methods based on artificial neural networks (AI) with representation learning.

Using **Deep Learning** models, we can use big volumes of data, including images and output a LiDAR image that **predicts the total biomass per patch of forest over a year.**

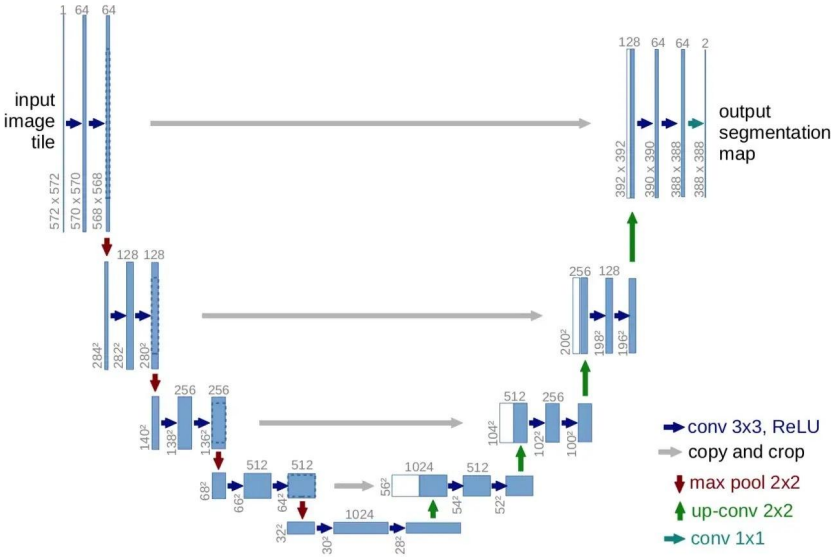
The algorithm could be extended to many different types of forest biomass, from boreal to temperate, and from tropical to arid.

Our Model.

We used convolutional neural networks already designed for image recognition: **Unet**.

With over **8 million parameters**, our model has been perfected to analyse satellite forest images, with **over 10 hours of training**.

The performance of our model is measured by **root mean squared error**.



Future considerations

After taking a deeper look at the dataset, plenty of the images from satellite 2 were **fully clouded**, and that **affects our predictions dramatically**.

Were we to have more time, perhaps we could have tested different models to take different considerations into account.

However... these models would have taken **68 hours to train** on the whole dataset.

Processing capacity and time were limiting factors.

The Prediction

BioMassters_Predictor

Thank you for your time !