University of Reading

Department of Computer Science

Generating Tree Species Classification and Change Maps to Assist Mitigate Climate Change

Pedro Junio

Supervisor: Muhammad Shahzad

A report submitted in partial fulfilment of the requirements of
the University of Reading for the degree of
Master of Science in *Data Science and Advanced Computing*

July 28, 2024

Declaration

I, Pedro Junio, of the Department of Computer Science, University of Reading, confirm that this is my own work and figures, tables, equations, code snippets, artworks, and illustrations in this report are original and have not been taken from any other person's work, except where the works of others have been explicitly acknowledged, quoted, and referenced. I understand that if failing to do so will be considered a case of plagiarism. Plagiarism is a form of academic misconduct and will be penalised accordingly.

I give consent to a copy of my report being shared with future students as an exemplar.

I give consent for my work to be made available more widely to members of UoR and public with interest in teaching, learning and research.

Pedro Junio
July 28, 2024

Abstract

TODO update at the end

Variety of tree species are crucial in reducing the vulnerabilities and offering stable ecosystem functioning. The precise quantification and assessment of existing tree species on global-scale is therefore essential in filling the science-policy gaps by providing key insights essential in promoting the success of natural climate solutions and devising effective climate mitigation policies. To this end, this study aims to develop novel deep learning based algorithms by using the multi-temporal multi-spectral imagery to generate large-scale forest/tree species classification maps.

Keywords: Deep Learning, Sentinel-2, Copernicus, Regression, Forests, Trees, Climate

Report's total word count: over 9000

Contents

1 Data Exploration **1**
1.1 EU-Forest Labels 1
1.2 Sentinel-2 Features 3

List of Figures

1.1 Map of the most common tree genera in EU-Forest. 1
1.2 Distribution of genera (left) and species (right) in EU-Forest. 2
1.3 Distribution of genera (left) and species (right) per location. 2
1.4 Sentinel-2 mission infographic. It highlights important facts and achievements of the mission. Courtesy of ESA. 3
1.5 Multiple sample locations overlaid on Google Earth (left) and Sentinel-2 (right) images. 4
1.6 Sample location overlaid on Google Earth (left) and Sentinel-2 (right) images. 4

List of Tables

Chapter 1

Data Exploration

1.1 EU-Forest Labels

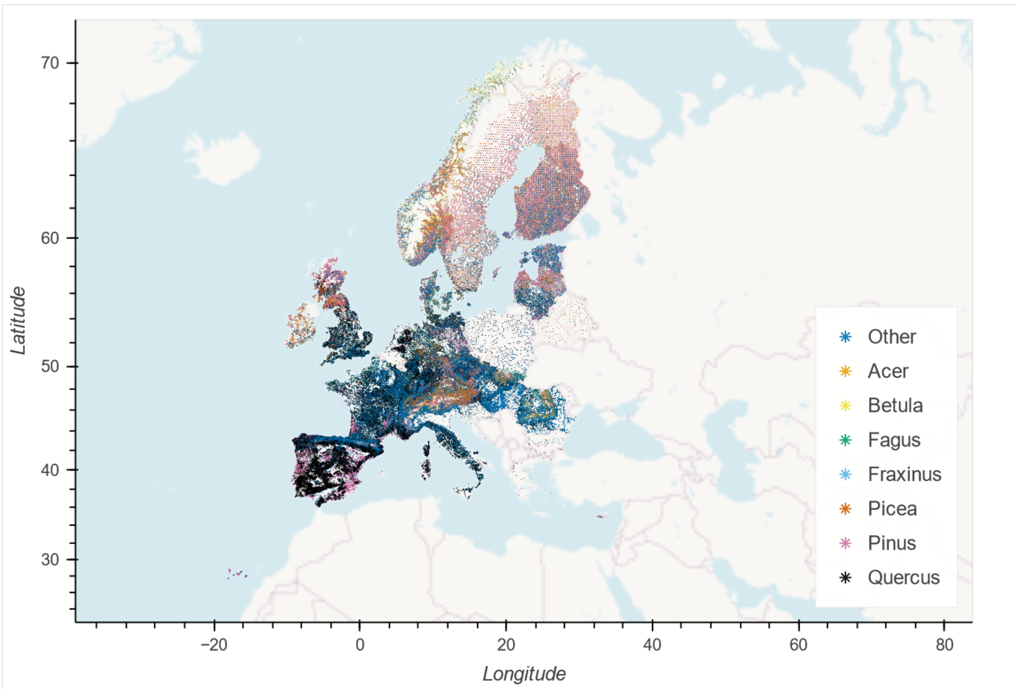

Figure 1.1: Map of the most common tree genera in EU-Forest.

EU-Forest is a dataset containing tree species and genera for nearly $250,000$ plots across Europe Mauri et al. (2017). Each plot is $1\,\text{km} \times 1\,\text{km}$ and may contain multiple tree species and genera. Fig. 1.1 shows the distribution of tree genera in the EU-Forest dataset across 21 European countries. In this figure, the label 'Other' is an umbrella class for 70 tree genera with less than 20,000 occurrences each.

Using the EU-Forest dataset to train a convolutional classifier of tree genera with Sentinel-2 data offers several significant advantages. Firstly, the dataset's high spatial resolution enables fine-grained analysis of tree species distribution, which enhances the accuracy of the classifier. Its comprehensive coverage across Europe, including diverse forest types and geographical areas, allows the model to learn from a wide variety of environments and tree genera. The rich occurrence data provides detailed information on tree species, aiding in precise identification and classification.

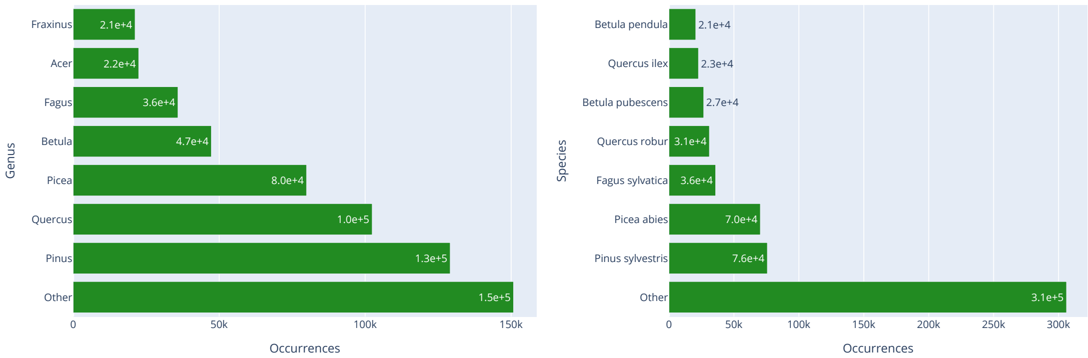

Figure 1.2: Distribution of genera (left) and species (right) in EU-Forest.

Integration with Sentinel-2 satellite data, which provides high-resolution multispectral images, allows for a robust model that leverages both ground-truth data and spectral information. The dataset, being relatively recent, offers a contemporary snapshot of forest conditions, ensuring that the trained model is relevant to current ecological and environmental conditions.

The plots in Fig. 1.2 underscore the prevalence of certain genera and species in European forests, providing valuable insights for training a convolutional classifier. The dominance of specific genera and species in the dataset can enhance the classifier's ability to accurately identify and classify tree types when combined with Sentinel-2 data.

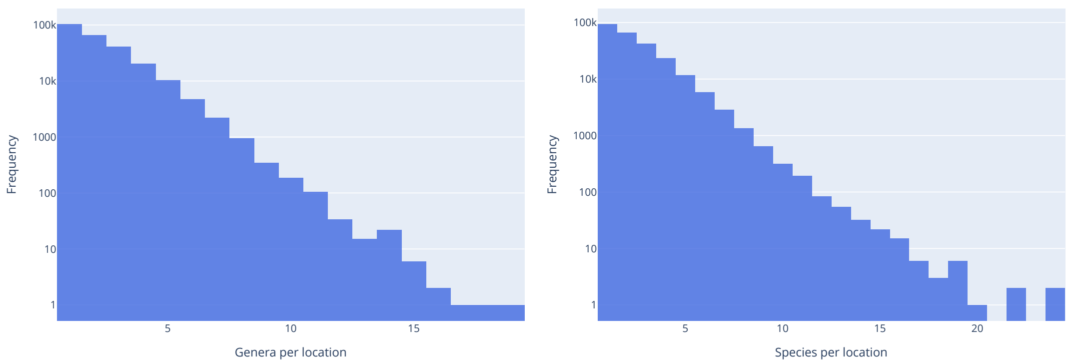

Figure 1.3: Distribution of genera (left) and species (right) per location.

The plots in Fig. 1.3 reveal a common pattern in biodiversity studies: most locations are characterized by a limited number of dominant genera and species, with a smaller number of locations exhibiting higher diversity. This pattern is important for training a convolutional classifier, as it indicates that the classifier will often encounter locations with limited genera and species. However, it must also be capable of handling the less common, more diverse locations. The high-frequency, low-diversity areas will likely dominate the training process, influencing the classifier's ability to generalise across different forest types.

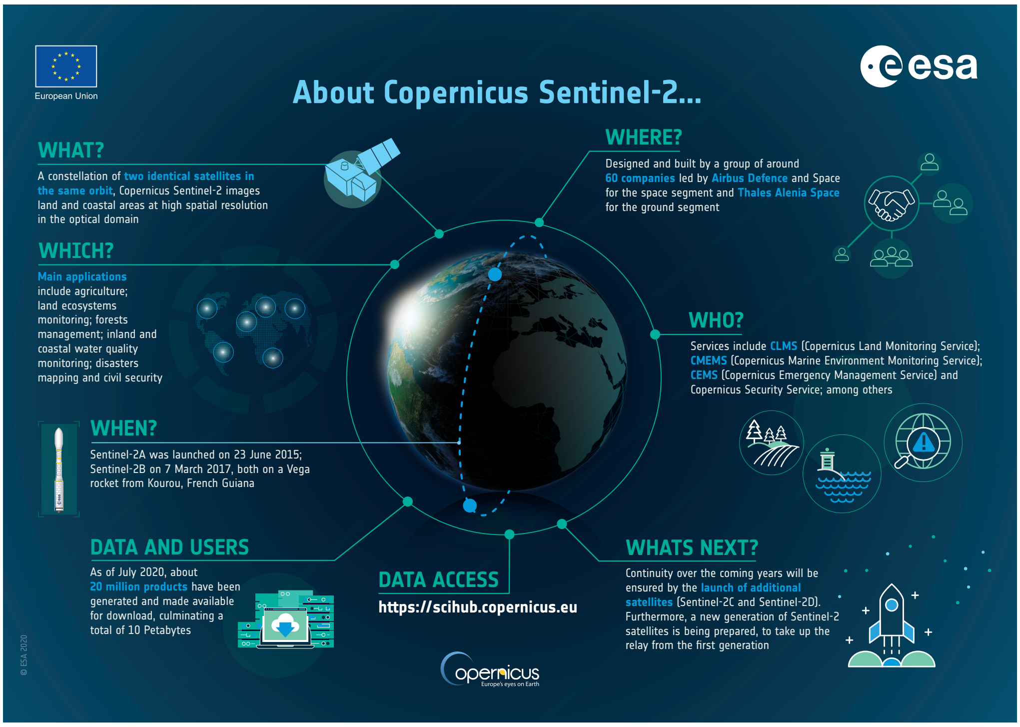

Figure 1.4: Sentinel-2 mission infographic. It highlights important facts and achievements of the mission. Courtesy of ESA.

1.2 Sentinel-2 Features

Using Sentinel-2, Fig. 1.4, specifically the 10-meter and 20-meter resolution bands, for training a convolutional classifier of tree genera offers several significant advantages. Sentinel-2 provides high-resolution imagery with these bands capturing detailed spatial information essential for precise classification tasks.

The 10-meter resolution bands include the visible (red, green, blue) and near-infrared (NIR) wavelengths, which are crucial for identifying vegetation health and differentiating tree genera based on their reflectance properties. The 20-meter resolution bands cover the red-edge, shortwave infrared (SWIR), and additional near-infrared regions, further enhancing the classifier's ability to distinguish between different tree genera by capturing subtle differences in spectral signatures.

The multispectral imaging capability of Sentinel-2, with these selected bands, allows for detailed analysis and precise classification of tree genera. Each genus reflects and absorbs light differently across these wavelengths, providing rich data for the classifier to learn from and accurately identify tree types.

Moreover, Sentinel-2 has a frequent revisit time, with satellites passing over the same area every 5 days at the equator. This frequent update cycle is crucial for handling cloud cover, as it increases the likelihood of acquiring cloud-free images, ensuring that the classifier is trained on clear and usable data.

Sentinel-2 also offers extensive geographical coverage, capturing large areas in each image. This comprehensive coverage is essential for training classifiers intended for wide-ranging applications across different forest types and regions, and it supports the development of global models for tree genus classification.

Additionally, Sentinel-2 data is freely available through the European Space Agency's

Copernicus program and Google Earth Engine. This open access removes budget constraints, making high-quality satellite data accessible for various research and operational purposes.

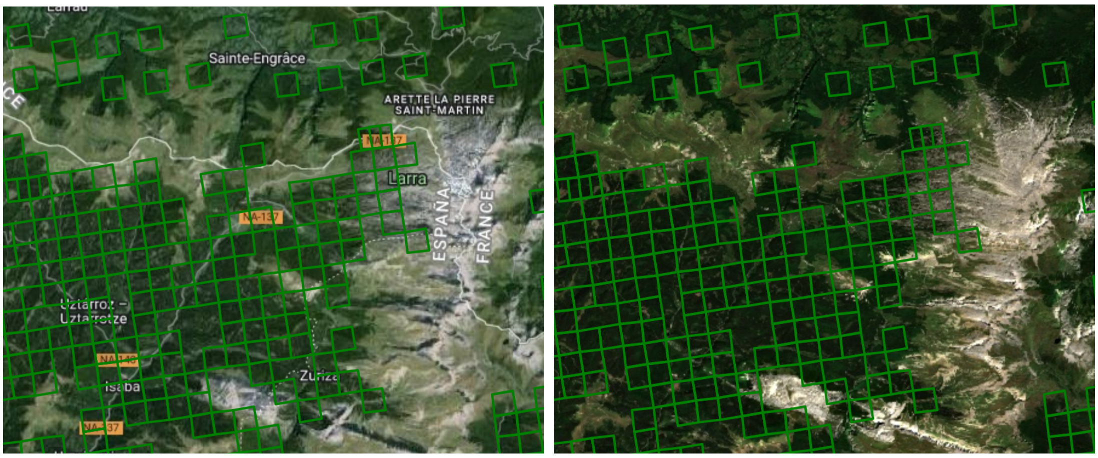

Figure 1.5: Multiple sample locations overlaid on Google Earth (left) and Sentinel-2 (right) images.

Figs. 1.5 and 1.6 illustrate the integration of high-resolution geographical context with Sentinel-2 satellite data for detailed environmental analysis. The grid overlay in the images represents areas for data collection and analysis. The left images provide a more detailed geographical context, while the right images show how Sentinel-2 bands B2, B3, and B4 (blue, green, and red respectively) are structured and utilised for spectral analysis.

Figure 1.6: Sample location overlaid on Google Earth (left) and Sentinel-2 (right) images.

References

Mauri, A., Strona, G. and San-Miguel-Ayanz, J. (2017), 'EU-Forest, a high-resolution tree occurrence dataset for Europe', *Scientific Data* **4**(1), 160123.