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 25, 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 25, 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

Chapter 1

Introduction

1.1 Background

Forests play a critical role in regulating the Earth's climate and supporting biodiversity. However, they are increasingly threatened by human activities and environmental changes, including deforestation, land-use conversion, and climate change. Understanding the composition and dynamics of forest ecosystems is essential for effective conservation and management efforts.

Accurately mapping tree species and monitoring changes in forest cover over time are fundamental tasks in ecosystem management and conservation. Such maps provide valuable information for assessing biodiversity, tracking habitat loss, and understanding the impacts of climate change on forest ecosystems.

This study aims to develop and evaluate methods for generating tree species predictions and change maps using remote sensing data. By leveraging advances in machine learning and image processing techniques, the project seeks to contribute to our understanding of how forests are responding to environmental changes and inform strategies for mitigating the effects of climate change.

1.2 Literature Review

Remote sensing technologies, including satellite imagery and LiDAR data, have revolutionised our ability to map and monitor forests at regional and global scales. Recent developments in machine learning algorithms, such as Convolutional Neural Networks (CNNs) and Random Forests (RF), have significantly improved the accuracy of tree species predictions from remotely sensed data (Bolyn et al. (2022), Ahlswede et al. (2023), Wessel et al. (2018)).

Detecting changes in forest compositions is essential for assessing the impacts of disturbances such as wildfire, climate change, and tree diseases (Krist et al. (2014)). Time-series analysis of satellite imagery, combined with advanced algorithms for change detection, can enable researchers to quantify forest dynamics and identify areas undergoing significant ecological transitions (Furusawa et al. (2023)).

Climate variables, including temperature, precipitation, and soil moisture, influence the distribution and health of tree species. Integrating climate data with remote sensing imagery allows researchers to model the relationships between environmental factors and vegetation dynamics, providing insights into how climate change is altering forest ecosystems (Mehmood et al. (2024)).

While progress has been made in remote sensing-based tree species classification and change mapping, several challenges remain. These include the need for improved methods for handling complex forest ecosystems, increase in the availability of ground-truth data, incorporating uncertainty into mapping algorithms, and scaling up analyses to cover larger geographical areas.

Chapter 2

Research Framework

2.1 Research Question

Can we generate tree species predictions and change maps using remote sensing data and existing ground-truth data to assist in mitigating the impacts of climate change on forest ecosystems?

2.2 Justification

Forests are vital for maintaining biodiversity, regulating the climate, and providing essential ecosystem services. However, they are increasingly vulnerable to the effects of climate change, including shifts in temperature and precipitation patterns, increased frequency and intensity of extreme weather events, and altered disturbance regimes. Understanding how forests are responding to these changes is crucial for effective conservation and management efforts.

By developing methods to generate tree species predictions and change maps from remote sensing data, this research will attempt to address a pressing need for tools to monitor and assess the impacts of climate change on forest ecosystems. Such maps can provide valuable insights into changes in species composition, distribution, and health over time, helping to identify areas of conservation concern, prioritise management interventions, and inform climate adaptation strategies.

2.3 Research Objectives

- Establish accurate prediction models using remote sensing data and ground-truth labels.
- Identify changes in forest cover over time, contingent on the success of the tree species prediction models.
- Analyse the influence of climate variables on forest dynamics, depending on the quality of tree species predictions and change detection results.
- Summarise findings and propose further research or practical applications.

2.4 Measurement of Achievements

- Evaluate the accuracy and reliability of tree species prediction and change detection algorithms.

- Assess the ability to effectively integrate climate data and analyse its influence on forest dynamics, conditional on the quality of initial results.
- Determine how useful the generated maps are for informing conservation and management decisions.
- Suggest next steps and areas for further investigation based on the study's findings.

Chapter 3

Methodology

3.1 Develop Methods for Tree Species Predictions

- **Data collection:** gather multi-temporal, multi-spectral satellite imagery covering the study area. Obtain ground-truth data on tree species composition through existing forest inventory datasets.
- **Data exploration:** visualise the data and use statistical techniques to describe dataset characterisations, such as size, quantity, and accuracy, in order to better understand the nature of the data.
- **Pre-processing:** pre-process the satellite imagery to remove noise, correct for atmospheric effects, and normalise.
- **Model training:** train machine learning algorithms, such as Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM), using the pre-processed data and ground-truth labels to develop accurate prediction models.

3.2 Detect Changes in Forest Cover

- **Time-series analysis:** obtain multi-temporal satellite imagery covering multiple time periods to capture changes in forest cover over time.
- **Change detection algorithm:** if feasible, implement change detection algorithms to identify areas of forest cover change between consecutive time steps.

3.3 Integrate Climate Data

- **Climate data acquisition:** subject to the quality of previous results, obtain climate data, such as temperature, precipitation, and soil moisture, from gridded climate datasets covering the study area or global averages.
- **statistical analysis:** perform statistical analysis to assess the relationships between climate variables and tree species dynamics, using techniques such as correlation analysis or regression modelling.

Chapter 4

Conclusions and Future Work

4.1 Conclusions

TODO

4.2 Future work

Efficiencies:

- band analysis, e.g. using correlations as the basis could perhaps reduce the model to around 4 Sentinel-2 bands
- anomalies, e.g. long tails after z-score
- avoid upsampling by using functional layering with multi-inputs, could save 50% storage and processing, e.g. could increase model depth
- consider the shading method, although a shard per sample simplifies the data pipeline, it could present challenges to a system with limited CPUs or slower storage devices such as magnetic disks

Climate:

- climate analysis currently uses 4 data points (2020, 2021, 2022, 2023), by reducing the date range of the training data to one year and by using next year's Sentinel-2 data, 7 data points would be available to perhaps provide a more statistically significant analysis

Data methods:

- the class imbalance reflects reality in that European forests are dominated by certain tree genera, currently a simple class weight is applied but there exist other robust methods such as over-sampling the minority classes
- the cut-off of 20,000 genus samples marks a threshold where a recall of 0.7 or above could be obtained, but in the process, dozens of genera were neglected, whereas EU member states have a total of 1.8 million plots (1 km x 1 km) Commission (2019), which represents around 7 times more samples than used here

Labels:

- forest fires have not been considered despite being a significant concern in climate change, the analysis could be re-run by first excluding samples within regions known to have had forest fires
- commercial vs non-commercial forests, EU member states have 182 million hectares of forest or woodlands, 19 million of which are forests in nature protection areas and could present a better target for understanding climate change effects due to conservation efforts
- species were grouped by genus due to data limitations, consider whether using species rather than genus could benefit the analysis

Analysis:

- the intent of this research is to facilitate a global analysis, a logical next step would be to add North American (or a similarly well-documented region) labels either by simply combining with existing labels or creating a separate model for each region

Summarising: Two initial choices for training the model: use the median of two periods in a year, one that captures warmer months and one that captures colder months, or use only warmer months. Avoid upsampling the lower resolution bands in order to nearly half data storage needs. Upsampling could be done later or even in the preprocessing pipeline using multiple methods if required. Applying these efficiency improvements could vastly increase the reach of this analysis to other regions of the world.

References

Ahlswede, S., Schulz, C., Gava, C., Helber, P., Bischke, B., Förster, M., Arias, F., Hees, J., Demir, B. and Kleinschmit, B. (2023), '*TreeSatAI Benchmark Archive*: a multi-sensor, multi-label dataset for tree species classification in remote sensing', *Earth System Science Data* **15**(2), 681–695.
URL: *https://essd.copernicus.org/articles/15/681/2023/*

Bolyn, C., Lejeune, P., Michez, A. and Latte, N. (2022), 'Mapping tree species proportions from satellite imagery using spectral–spatial deep learning', *Remote Sensing of Environment* **280**, 113205.
URL: *https://www.sciencedirect.com/science/article/pii/S0034425722003145*

Commission, E. (2019), 'Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions stepping up eu action to protect and restore the world's forests', *Committee on the Environment, Public Health and Consumer Policy* .

Furusawa, T., Koera, T., Siburian, R., Wicaksono, A., Matsudaira, K. and Ishioka, Y. (2023), 'Time-series analysis of satellite imagery for detecting vegetation cover changes in indonesia', *Scientific Reports* **13**(1).

Krist, F. J., Ellenwood, J. R., Woods, M. E., McMahan, A. J., Cowardin, J. P., Ryerson, D. E., Sapio, F. J., Zweifler, M. O. and Romero, S. A. (2014), '2013-2027 national insect and disease forest risk assessment.', *Fort Collins, CO: US Forest Service, Forest Health Technology and Enterprise Team* .

Mehmood, K., Anees, S. A., Muhammad, S., Hussain, K., Shahzad, F., Liu, Q., Ansari, M. J., Alharbi, S. A. and Khan, W. R. (2024), 'Analyzing vegetation health dynamics across seasons and regions through ndvi and climatic variables', *Scientific Reports* **14**(1), 11775.

Wessel, M., Brandmeier, M. and Tiede, D. (2018), 'Evaluation of different machine learning algorithms for scalable classification of tree types and tree species based on sentinel-2 data', *Remote Sensing* **10**(9).
URL: *https://www.mdpi.com/2072-4292/10/9/1419*