Masters Program in Geospatial Technologies

ABOVE GROUND BIOMASS AND CARBON SEQUESTRATION ESTIMATION
Implementation of a Sentinel-2 based exploratory workflow.

Giulia Molisse

Dissertation submitted in partial fulfilment of the requirements
for the Degree of *Master of Science in Geospatial Technologies*

NOVA IMS Information Management School
Universitat Jaume I
Westfälische Wilhelms-Universität Münster

Above Ground Biomass and Carbon Sequestration Estimation

Implementation of a Sentinel-2 based exploratory workflow

Dissertation supervised by

Hugo Costa
Dzhaner Emin
Filiberto Pla Bañón

February 2021

Declaration of originality

I declare that the work described in this document is my own and not from someone else. All the assistance I have received from other people is duly acknowledged and all the sources (published or not published) are referenced. This work has not been previously evaluated or submitted to NOVA Information Management School or elsewhere.

Lisbon, February 2021

Giulia Molisse

Acknowledgements

*Here **we** are.*

For funding my postgraduate studies, thanks to the existence of merit-based scholarships. For investing in me and funding this work, a grateful thanks to IABG.

A special thanks to Dzhaner Emin. He supervised my Master thesis with professionalism, interest and - as if that wasn't enough - with a great deal of tips on academic writing. I would also like to acknowledge the effort of my supervisors Prof. Hugo Costa and Prof. Filiberto Pla Bañón. A sincere thanks to my Bachelor supervisor, Prof. Giacomo Certini, who left me with high expectations on what a Professor should be like; to Prof. Marco Painho, who met these expectations, and who constantly shows his students an enviable academic enthusiasm.

To the welcoming and kind colleagues at IABG, who made my months there extremely valuable. To each and every one of my University colleagues, now dear friends. Despite this program was remarkable, you have been - hands down - my favourite part of it. I wish I could mention you all.

*And now, **my dearest**.*

To my family. I am here because of my incredible parents, who worked hard to provide me and my brother Lorenzo with a good life; and had yet to deal with our moody teenage years. I hope this will make up for that. To Lorenzo, you are the next one. To António, my dear boyfriend and the most supportive man on the planet.

To Sole and our talks at dusk, I am proud of who you are. A big thanks to Erika, Marina, Emili, Arianna, Lorenzino, Lucrezia and Benedetta, for their invaluable and warm friendship.

Finally, and with all my heart, I want to thank my little cousins: Marta, Pietro, Tecla, Irene, Allegra, Gabriel and Amelie. Long list, but I want them to see their names here. You see? University is not a big deal. If I could do it, you can too!

Above Ground Biomass and Carbon Sequestration Estimation

Implementation of a Sentinel-2 based exploratory workflow

Abstract

This work presents a Sentinel-2 based exploratory workflow for the estimation of Above Ground Biomass (AGB) and Carbon Sequestration (CS) in a subtropical forest. In the last decades, remote sensing-based studies on AGB have been widely investigated alongside with a variety of sensors, features and Machine Learning (ML) algorithms. Up-to-date and reliable mapping of such measures have been increasingly required by international commitments under the climate convention as well as by sustainable forest management practices.

The proposed approach consists of 5 major steps: 1) generation of several Vegetation Indices (VI), biophysical parameters and texture measures; 2) feature selection with Mean Decrease in Impurity (MDI), Mean Decrease in Accuracy (MDA), L1 Regularization (LASSO), and Principal Component Analysis (PCA); 3) feature selection testing with k-Nearest Neighbour (kNN), Random Forest (RF), Extreme Gradient Boosting (XGB), and Artificial Neural Network (ANN); 4) hyper-parameters fine-tuning with Grid Search, Random Search and Bayesian Optimization; and finally, 5) model explanation with the SHapley Additive exPlanations (SHAP) package, which to this day has not been investigated in the context of AGB mapping.

The following results were obtained: 1) MDI was chosen as the best performing feature selection method by the XGB and the Deep Neural Network (DNN), MDA was chosen by the RF and the kNN, while LASSO was chosen by the Shallow Neural Network (SNN) and the Linear Neural Network (LNN); 2) before hyper-parameters optimization, the Deep Neural Network (DNN) yielded the best performance with a Root Mean Squared Error (RMSE) of 42.30 t/ha; 3) after hyper-parameters fine-tuning with Bayesian Optimization, the XGB model yielded the best performance with a RMSE of 37.79 t/ha; 4) model explanation with SHAP allowed for a deeper understanding of the features impact on the model predictions. Finally, the predicted AGB throughout the study area showed an average value of 83 t/ha, ranging from 0 t/ha to 346.56 t/ha. The related CS was estimated by using a conversion factor of 0.47.

All the illustrations were done by the author

Keywords

Above Ground Biomass
Carbon Sequestration
Medium resolution
Random Forest
Extreme Gradient Boosting
Artificial Neural Network
k-Nearest Neighbour
Feature Selection
Bayesian Search
Sentinel-2

Index of the text

Index of Figures

Index of Tables

List of Abbreviations

AGB Above Ground Biomass

ANN Artificial Neural Network

BGB Below Ground Biomass

CGAT Geo-Environmental Cartography and Remote Sensing Group

CHM Canopy Height Model

Clre Chlorophyll index from red-edge bands

CS Carbon Sequestration

DBH Diameter at the Breast Height

DEM Digital Elevation Model

DNN Deep Neural Network

EO Earth Observation

EU European Union

FAO Food and Agricultural Organization

FAPAR Fraction of Absorbed Photo-synthetically Active Radiation

FCOVER Fraction of vegetation cover

GEMI Global Environmental Monitoring Index

GIS Geographic Information System

GLCM Gray Level Co-occurrence Matrix

GNDVI Green Normalized Difference Vegetation Index

H Height

IPCC Intergovernmental Panel on Climate Change

kNN k-Nearest Neighbours

LAIcb Canopy Chlorophyll Content

LAIcw Canopy Water Content

LAI Leaf Area Index

LASSO L1 Regularization

LNN Linear Neural Network

MDA Mean Decrease in Accuracy

MDI Mean Decrease in Impurity

ML Machine Learning

MSAVI Modified Soil Adjusted Vegetation Index

NDVIre Normalized Difference Vegetation Index based on red-edge bands

NDVI Normalized Difference Vegetation Index

NDWI Normalized Difference Water Index

NOAA National Oceanic and Atmospheric Administration

PCA Principal Component Analysis

RF Random Forest

RMSE Root Mean Squared Error

RQ Research Question

SAVI Soil Adjusted Vegetation Index

SHAP SHapley Additive exPlanations

SNN Shallow Neural Network

SOM Soil Organic Matter

UNFCCC United Nations Framework Convention on Climate Change

UPV Universitat Politècnica de València

VI Vegetation Index

XGB Extreme Gradient Boosting

Chapter 1

Introduction

The following thesis was funded by the MAIL project [1]*, a framework aiming to identifying marginal lands in Europe and strengthening their contribution potentialities in a Carbon sequestration strategy. The entire project involves a total of 6 academic and non academic partners* [2] *from 4 European Member States. Furthermore, the project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement. This thesis and its findings represent part of the work being developed for the MAIL project.*

International commitments under the climate convention (UNFCCC, 1997) as well as sustainable forest management practices, require accurate and up-to-date mapping of forested areas. Forests are the largest terrestrial carbon pool, counting about 85% of the total land vegetation biomass, and in them it is found 73% of the global soil carbon (Schimel et al., 2001). Additionally, global warming reports require detailed information on the forests' carbon content (IPCC, 2006). Hence, reliable and up-to-date Above Ground Biomass (AGB) and Carbon Sequestration (CS) mapping has become worldwide essential.

Within the terrestrial pool, CS can be defined as the sum of carbon stored in vegetation biomass and soil, and it is often expressed in tonnes over hectares (t/ha) (Salem et al., 2020). Nonetheless, this work focuses exclusively on the portion of carbon stored in vegetation biomass, more exactly on the above ground portion of a forested area, which includes stems, stumps, branches, bark, seeds, foliage etc., also known as AGB and expressed in tonnes over hectares (t/ha). The decision of predicting only the above-ground portion of the forested area is justified by the fact that, while on one hand the Below Ground Biomass (BGB) can be estimated to be around 20% of the predicted AGB, on the other hand, the estimation of AGB itself is not straightforward, as it requires reliable field measures, images, and algorithms. Moreover, the majority of the remote-sensing based studies on biomass estimation focus on AGB and the related CS.

The encounter of Earth Observation (EO) technologies with AGB and CS mapping brought multiple advantages to this field. Traditional forest inventory methods - i.e. field plots surveys - are expensive, time-consuming, and can demand for the destruction

[1] `https://marginallands.eu/`

[2] Space Research Centre of the Polish Academy of Sciences, AUTH, HOMEOTECH, Universitat Politècnica de València, IABG, and CESEFOR.

of portion of the forest (Salem et al., 2020). Remote sensing allows for a time-effective, affordable and, constant monitoring of the state of forests. Moreover, these advantages enable for a larger scale mapping compared to the sole use of traditional methods. Estimation of AGB has been widely explored in the context of EO and GIS technologies (Zheng et al., 2004; Pandit et al., 2018; Gao et al., 2018; Cairns et al., 1997). Remote sensing-based estimations of AGB utilise a variety of sensors, features, methods to select the optimal features combination, as well as several regression models (Galidaki et al., 2017; Lu, 2006). In a literature review carried on by Salem et al. (2020), over 150 peer-reviewed articles on AGB and CS were analysed; their review highlights how these studies tend to follow a similar recurrent workflow, which can be parted into the following major steps: collection of forestry's inventory measures, of images, as well as of GIS-based miscellany data, preprocessing, features generation and selection, regression algorithms training and comparison, and production of the final maps. Up until now, many studies have tested the use of different sensors, predictors, and regression models; the ones which were considered relevant for this work are investigated in more detail in Chapter 2.

This work aims to asses the capabilities of Sentinel-2 derived measures for the estimation AGB in a subtropical forest. More specifically, the author aims to identify an optimal Season for satellite data acquisition, a ranking of the most influential satellite-generated features, and to build an optimal regression model. Based on these three objectives, the following Research Questions (RQs) were defined:

1. Feature selection is a widely known important step in helping against the risk of model over-fitting and long training time. Hence, the most representative features are going to be identified and used as model inputs by answering to the following question: aiming to the identification of an optimal number and type of features, which is the best performing feature selection method when predicting AGB in the study area?

2. Highly connected to the previous RQ, this work seeks to understand which are the most influential features - i.e. season, Vegetation Indices (VIs), texture measures, etc. - when predicting AGB?

3. Machine Learning (ML) regression algorithms can be characterised by many hyper-parameters which require to be fine-tuned in order to improve the model predictions. Manual hyper-parameter fine-tuning is time consuming and non-reproducible, thus, the following question was posed: which hyper-parameter fine-tuning method yields the best performance for the chosen regression models?

4. In terms of Root Mean Squared Error (RMSE), which ML regression algorithm produces the most accurate estimate?

5. Finally, other than evaluating the model solely based on its error, a deeper understanding of how the selected features influence the model predictions can bring valuable insights in the field of AGB and CS mapping. This is achieved by answering to the following question: how do the top-ranked features and their interaction impact on the model output?

With the intention of answering the stated RQs, our methodology is going to broadly

follow the aforementioned commonly implemented steps. Additionally, touches of novelty will be introduced throughout the entire workflow.

In order to select which features - i.e. VIs, texture measures, etc. - allow for a better performing model, features selection is often implemented; previous studies on AGB tend to follow one of these approaches: step-wise regression analysis, a linear method, or measures obtained from the Random Forest (RF) algorithm (Gao et al., 2018; Lu, 2006). Hence, with the purpose of answering the 1st and 2nd RQs, 4 features selection techniques - Mean Decrease in Impurity (MDI), Mean Decrease in Accuracy (MDA), L1 Regularization (LASSO), and Principal Component Analysis (PCA) - are tested and the features rankings analysed. Subsequently, the comparison among features selection methods is performed by implementing commonly used regression algorithms in this field - RF, K-Nearest Neighbour (kNN) and Artificial Neural Networks (ANN). Furthermore, we introduce and compare the Extreme Gradient Boosting (XGB) algorithm (Chen and Guestrin, 2016); which is amongst the most popular variants of the Gradient Boosting algorithm and it has been the winner of many Machine Learning (ML) competitions (Seif, 2019).

RF and XGB are characterised by several hyper-parameters which can be adjusted by the user. While some of these do not have a major impact of the model's predictive capacity, others, such as the number of estimators, can quite influence the predictions. In order to answer the 3rd and 4th RQs, this work aims to make the most of such a characteristic by exploring and comparing 3 types of hyper-parameters searches - Grid, Random, and Bayesian search.

Additionally, AGB will be mapped by using the predictions from the better performing feature selection method, regression model and hyper-parameter search method.

Finally, the last step of this workflow aims to answer the 5th and last RQ, which seeks to understand how the model output is influenced by the selected features. ML algorithms are often seen as black boxes, which refers to the difficulty in understanding how and to what extent the inputs and other components of a model impact on the final output (Dallas, 2017). Therefore, with the purpose of bringing the research surrounding AGB and CS a step forward, we are going to investigate how certain features influence the model output. To do this, a model explainer is introduced; these algorithms have the ability to explain the output of ML models; SHapley Additive exPlanations (SHAP), recently proposed by Lundberg et al. (2020), achieves that by showing how each features contributes to push the model output from a baseline - identified as the average model output over the training dataset - to the final output for each instance. This allows to assess the influence that the inputs have in the predictions, which can be of interest for future studies, especially when dealing with similar environmental conditions. To our best knowledge and up to this day, the SHAP package was never implemented in the context of Above Ground Biomass (AGB) mapping.

Chapter 2

Background & Related Work

Carbon Sequestration (CS) is a natural process involving the capture of carbon dioxide from the atmosphere and its long-term storage into 3 major carbon pools; namely, terrestrial, oceanic, and geological pools (IPCC, 2006; Salem et al., 2020). Furthermore, carbon can be stored in either a liquid or a solid state by different means such as trees, soil, ocean or organic matter (Lackner, 2003).

The Intergovernmental Panel on Climate Change (IPCC, 2006) divides the terrestrial pool into 5 main reservoirs: Above Ground Biomass (AGB), Below Ground Biomass (BGB), litter, woody debris, and Soil Organic Matter (SOM).

2.1 Above Ground Biomass and Current Carbon Sequestration

Carbon Sequestration (CS) in the terrestrial pool is the sum of the amounts of carbon in vegetation biomass and soil (Salem et al., 2020). Accounting for largest portion of the the terrestrial vegetation biomass, forests have been widely studied; Indeed, forests are characterized by the highest carbon density compared to other terrestrial environments (Stinson et al., 2011), building up to circa 80% of the above-ground carbon stored in the terrestrial pool (IPCC, 2006). However, because forests are affected by many disturbances - i.e. fires, deforestation, parasites, land use change, etc. - around 60% of newly stored carbon is cyclically returned to the atmosphere (Salem et al., 2020). These characteristics make forest monitoring an important step towards up-to-date and reliable information about the state global carbon budget, particularly in the context of climate change mitigation (Galidaki et al., 2017).

Estimation of AGB has been widely explored in the context of remote sensing and GIS (Zheng et al., 2004; Pandit et al., 2018; Gao et al., 2018; Cairns et al., 1997). Previous studies estimate BGB and litter in mature forests to be around 20% and 10-20% of the predicted AGB, respectively (Cairns et al., 1997; Kankare et al., 2013).

If on one hand, BGB and litter can be estimated to be a fraction of AGB, on the other hand, providing a reliable estimation of the latter is not straightforward. Furthermore, the

estimation of CS from AGB is commonly performed by multiplying AGB by an average conversion factor of 0.5, which assumes the 50% of the dry biomass to be carbon (Khan et al., 2020). More precise factors have been provided for different ecological zones; the IPCC (2006) suggests a factor of 0.47 for tropical and subtropical forests. However, when the estimation of CS requires for higher precision, it should be noticed that the carbon content in dry matter changes between tree species, as well as amongst climate zones, and it is correlated to the species wood density; species with higher wood density store higher quantities of carbon; that is Mediterranean species have higher carbon content than tropical species, despite their extent and density being more limited (Thomas and Martin, 2012)

2.2 Above Ground Biomass estimation

AGB estimation has been the subject of extensive research, given its importance in planning carbon emission mitigation strategies and sustainable forest management. The estimation of AGB can be performed through direct or indirect methods (Salem et al., 2020; Galidaki et al., 2017; Lu, 2006)

Direct methods, also known as destructive, call for the harvesting, separation into components, oven drying, and weighing of the tree components as fractions of the biomass. These methods are highly precise, however, they are also expensive, destructive, and time-consuming, which makes them unsustainable to be used in large scale (Salem et al., 2020; Picard et al., 2012). The outputs from direct methods have been used through the years for building allometric equations - i.e. statistical models belonging to the indirect methods category.

Indirect methods, also known as non-destructive, do not require the physical destruction of trees. These can be further divided into two main approaches: allometric equations and Remote Sensing-based estimations (Galidaki et al., 2017), the latter being dependent on the first one for model training.

Allometric equations are statistical models that use forest measurable biophysical characteristics - i.e. Height (H), Diameter at Breast Height (DBH) or crown size - in order to estimate either tree volume or biomass; such models allow for the estimation of biomass without the need of harvesting (Vashum, 2012). Allometric equations have been developed for many tree species as well as for the most common species combination (Salem et al., 2020).

2.3 Remote sensing and GIS-based Above Ground Biomass estimation

Remote sensing-based estimation of AGB utilizes a variety of sensors, features and several regression models (Salem et al., 2020; Galidaki et al., 2017; Lu, 2006). Furthermore, another recurrent and important aspect to be considered when developing such methodologies is the ecological zone in which the forest is located, as each zone differs in terms of forest carbon content (Liu et al., 2012).

In their review with a focus on Mediterranean areas, Galidaki et al. (2017), assessed that, in terms of climate zones, AGB estimation has been extensively studied in boreal and tropical environments; this might be explained by the wide extent of tropical forests. Whereas, there is a limited amount of research on forests located in subtropical zone, especially Mediterranean forests. Despite these forests cover barely 85 million hectares (ha), representing only 2% of the world's forest area (FAO, 2012), the mean wood carbon content in the Mediterranean region, as well as in boreal and temperate regions, has been measured to be higher the tropical forests content (Thomas and Martin, 2012), between 45.7% and 60.7%; this makes Mediterranean forests valuable carbon pools and interesting areas to monitor.

Consequently, AGB estimation in tropical (Hernandez-Stefanoni et al., 2020; Nuthammachot et al., 2020; Wani et al., 2015; Vicharnakorn et al., 2014), subtropical (Pandit et al., 2018; Gao et al., 2018), temperate (Khan et al., 2020; Coulibaly et al., 2008; Zheng et al., 2004) and boreal (Puliti et al., 2020) forests is expected to result in different ranges of biomass and carbon content.

Sensors

It their extensive literature review, Salem et al. (2020), analysed over 150 peer-reviewed articled on AGB and CS estimation. Only 25% of these studies use imagery data from active sensors, almost equally divided into LiDAR (laser imaging, detection, and ranging) and Radar (radio detection and ranging) technologies; while the majority of the them deals with imagery collected from various passive sensors. Hereafter, passive sensors are going to be divided based on their spatial resolution into 3 categories: coarse ($>$ 100 m), moderate (10 - 100 m), and fine (sub-meter to 5 m) resolution.

Coarse resolution images, such as MODIS, SPOT Vegetation, and National Oceanic and Atmospheric Administration (NOAA), are valuable products when working at the continental or global extent as they provide consistent spatial and temporal data, which increases the probability to collect cloud-free images. However, these images perform poorly in discriminating the vegetation structure and data saturation occurs in high density forests (Salem et al., 2020); furthermore, mixed pixels can be an issue at this resolution, as there is no accordance between the size of the pixels and of the forestry inventory field plots.

SPOT, Landsat and Sentinel-2 provide medium spatial resolution images, these have been often used for AGB estimation (Zheng et al., 2004; Vicharnakorn et al., 2014; Wani et al., 2015). Zheng et al. (2004) use DBH together with various Vegetation Indices (VI) derived from Landsat 7 ETM+ to produce an initial biomass map in a temperate forest. This first map, together with land cover information, was then used to retrieve forest age as a further parameter in order to improve the predictions. Landsat TM and 5 images have been explored in tropical forests (Vicharnakorn et al., 2014; Wani et al., 2015) to extract VIs to be used as inputs in the regression model together with a Digital Elevation Model (DEM).

Zhao et al. (2016) proved that, when working with Landsat images, the performance of the regression models starts decreasing as the biomass density surpasses a threshold value of 130 t/ha; this value decreases as the forest structure becomes more heterogeneous. On the other hand, the temporal continuity of the Landsat mission provides with valuable data, especially when the focus resides on change detection at regional extents and on monitoring vegetation over long periods of time.

Sentinel-2A was launched in 2015 by the European Space Agency (ESA) and since then its potential in AGB estimation has been evaluated by several authors and within a variety of ecological zones (Pandit et al., 2018; Khan et al., 2020; Puliti et al., 2020).

The work of Pandit et al. (2018) explores the performance of spectrally derived indices from Sentinel-2A as inputs in a Random Forest (RF) model in a subtropical forest in Nepal. Field-based AGB values were estimated by applying an allometric equation using forestry inventory data from 113 measured plots with a radius of 12 m. The model performance was assessed by using a 10-fold cross-validation. The predicted AGB ranged from a minimum of 35.42 t/ha to a maximum of 276.92 t/ha, with a mean of 160 t/ha; the final model resulted in a Root Mean Squared Error (RMSE) of 25.32 t/ha between the observed and the predicted biomass.

Khan et al. (2020) explored the use of Sentinel-2 images in a mountainous temperate forests in Pakistan. Their study examines the performance of 3 categories of spectrally derived VIs - i.e. Broadband, Canopy Water Content, and Narrow band red-edge VIs. Out of 25 indices, only 11 were used as inputs in the linear regression model, most of which were red-edge VIs. The predicted AGB ranged from a minimum of 46.45 t/ha to a maximum of 279.59 t/ha, and a mean of 148.79 t/ha. The final biomass map was validated using 10 out of the 55 plots, with a radius of 17 m, and resulted in a RMSE of 35.23 t/ha between observed and predicted biomass.

Additionally, Sentinel-2 single bands have been used in boreal areas for estimating AGB over the entire Norwegian territory. In their work, Puliti et al. (2020), used the whole Norwegian forestry inventory, composed of 7710 plots with a radius of 9 m. Furthermore, a Canopy Height Model (CHM) was included as model input; this was generated by normalising a freely available 2 m resolution DEM covering the entire area north of 60 degrees of latitude with a freely available 10 m resolution Digital Terrain Model (DTM). The final maps were evaluated using cross-validation, resulting in a a RMSE of 45.8 t/ha when using solely Sentinel-2 single bands; a similar performance was achieved from using solely the CHM, with a RMSE OF 47.7 t/ha; finally, a noticeable synergy was found when using both, with a decrease of the error down to 41.4 t/ha.

Studies involving the twin satellites Sentinel-2A and 2B show higher performance in AGB estimation compared to working with Landsat images. The better performance that characterizes Sentinel-2 can be attributed to its higher spectral, spatial and temporal resolution compared to the Landsat mission. Specially, the presence of red-edge bands makes these images highly valuable for vegetation analysis (Pandit et al., 2018).

Additionally, studies which use fine spatial resolution sensors such as Quickbird, WorldView-2, or IKONOS, have been reviewed by Salem et al. (2020); these images require for a large storage space and the high cost of image acquisition limites their use to small study areas. IKONOS, together with topographic, soil type, and precipitation data were utilized to map AGB in a temperate forest in Canada for 5 species groups (Coulibaly et al., 2008). For this work, texture measures and VIs were generated, and AGB estimation was computed using an Artificial Neural Network (ANN) and multiple regression models, as well as stratification based on species and Diameter at the Breast Height (DBH). The lowest RMSE (0.011 t/ha) was recorded by using the ANN for mapping spruce for the 9-18 cm DBH class.

The performance of Radar sensors in AGB estimation depends on which bands are installed on the sensor. L and P-bands are often used for predicting AGB as they showed to be highly correlated with such measure, while C-band resulted in little to none correlation with it (Salem et al., 2020). ALOS PALSAR is a L-band Synthetic Aperture Radar widely used in AGB mapping until 2011, when it stopped being operational. In a study from Hernandez-Stefanoni et al. (2020), ALOS PALSAR was integrated with LiDAR and climate data for estimating AGB in a tropical forest in Mexico. A total of 287 forestry inventory plots from 2009 and 2014 were corrected and used for model training and validation. Additionally, LiDAR was used to collect vegetation measures complementary to the National Forestry Inventory. Climatic water deficit, back-scatter and texture measures were used as inputs in RF models, by using different combination of training and validation samples. The residuals of the 4 models were further analysed showing no significant spatial auto-correlation. The average estimated AGB values was 99.2 t/ha with a standard deviation of 48.1 t/ha. Finally, the models were evaluated using cross-validation, which yield to a RMSE of 48.8 t/ha when modelling with field data, of 86.7 t/ha with LiDAR data, and of 57.1 t/ha with a combination of the two. Overall, L-Band encountered saturation issues in areas with an AGB values higher than 150 t/ha; however, the authors points out that this value is subject to change depending on vegetation type, complexity, or topography (Hernandez-Stefanoni et al., 2020), making radar sub-optimal for mountainous regions (Salem et al., 2020).

Other studies explored the combination of several datasets, such as Sentinel-1, Sentinel-2, ALOS PALSAR, and Landsat. Landsat TM, ALOS PALSAR and their combination was explored in a mountainous subtropical forest by Gao et al. (2018). In this study, regression models were tested and compared with and without stratification - i.e. performing the estimation at the species or forest type level - by utilizing a forest types map. Spectral signatures, back-scatter and texture measures were extracted from Landsat TM and ALOS PALSAR images and the selection of variables to input in the models was done using both step-wise regression and random forest measures. Finally, several models were tested, including linear regression and non-parametric algorithms. Under non-stratification conditions, the estimated AGB ranged from 25.7 to 180.7 t/ha, with a standard deviation of 33.9 t/ha. The results show a better performance of Landsat TM

with a RMSE of 27.7–29.3 t/ha compared to ALOS PALSAR which yielded a RMSE of 30.3–33.7 t/ha, while the combination of the two gave none to slightly improved performances depending on the model used. Moreover, overestimation and underestimation occurred independently on the dataset used. Gao et al. (2018) observe that for Landsat images, when AGB was within 50–130 t/ha, the residuals were relatively small, while for ALOS PALSAR this range was narrower, of about 90–120 t/ha. Such conclusion confirm that overestimation of small AGB values and underestimation of large ones are an issue for both optical and active sensors, such residuals increase as AGB values are higher than 150 t/ha or lower than 40 t/ha (Zhao et al., 2016; Gao et al., 2018)

Furthermore, a few studies explored the combination of Sentinel-2 and Sentinel-1 images estimating AGB (Nuthammachot et al., 2020; Forkuor et al., 2020). In a recent study carried out in an Indonesian tropical forest, AGB was modelled by 3 multiple linear regression models, using as inputs, Sentinel-2, Sentinel-1 and the combination of the two, and 45 forestry inventory plots (Nuthammachot et al., 2020). The models were then evaluated using 15 plots and showing RMSE of 37.35 t/ha for the model using solely Sentinel-1 back-scatter; Sentinel-2 derived indices shown higher performance, yielding to a RMSE of 19.08 t/ha; the poor performance is explained by the fact that Sentinel-1 works on the C-band. Finally, a combination of the two resulted in a slightly improved performance with a RMSE of 18.71 t/ha. The predicted average value from the best model was was $70.38t/ha$, with a minimum of $21t/ha$ and a maximum of $226t/ha$.

Another study carried out in a tropical forest in West Africa explored the potential of Sentinel-1 and 2 in AGB estimation (Forkuor et al., 2020). Spectral signature, biophysical parameters, VIs, Radar back-scatter and Radar difference and sum bands were used in different combination as inputs in RF models, together with 218 forestry plots. Validation was performed by using 40% of the forestry samples; modeling with Sentinel-2 achieved better results than using Sentinel-1 data, with a RMSE of 60.6 t/ha, and a RMSE of 78.6 t/ha, respectively. The best performance was reached by the combination of the two, with a RMSE 45.4 t/ha. Furthermore, modeling with solely Sentinel-2 spectral bands would sill achieve better results than using all Sentinel-1 generated features or subsets of it. The best model estimated AGB values ranging from 0.1 t/ha to 1042 t/ha, with 90% of the study area characterised by value of less than 90 t/ha.

Features

The choice of which and how many features to include when modelling a regression algorithm has an impact on the model performance. A summary of the commonly used features in AGB was done by Salem et al. (2020), and these are: spectral bands and derived measures such as VIs, biophysical parameters and texture; other ancillary data often used are topographic information, weather data, soil maps as well as land cover maps; another important factor is the season in which the images are collected. Furthermore, complex measures do not always perform as better predictors, this is the case in a study carried out in a tropical dry forest, where Sentinel-2 single bands resulted in more accurate outcomes compared to different combinations of SAR data (Forkuor et al., 2020).

Vegetation Indices (VIs) are mathematical transformations of the original spectral bands (Salem et al., 2020); different combinations perform better depending on certain forest characteristics - i.e. slope, density, complexity, etc. Khan et al. (2020), in his study for AGB mapping, divides VIs into 3 categories: Broadband, Canopy Water Content and Narrow red-edge band VIs.

Broadband VIs use the Near-infrared (NIR) band together with either the red or the green band; these VIs are sensitive to canopy leaf area, however they suffer from saturation problems as the forest density increases (Khan et al., 2020). The so called Canopy Water Content VIs utilise NIR and Short-wave infrared (SWIR) spectral bands; these VIs help understanding the forest health as vegetation in water stress has less carbon storage capacity compared to healthy individuals (Khan et al., 2020). Narrow red-edge band VIs utilise NIR together with Red-Edge bands; such indices are often used in vegetation studies as they detect photosynthesis in more detail and are less affected by saturation problems than broadband VIs (Khan et al., 2020; Pandit et al., 2018).

Biophysical parameters can also be detected through remotely-sensed images. These manifest themselves through roughness, shadow and spectral response, and are often used in vegetation studies as they help discriminating the spatial distribution and dynamics of vegetation (Forkuor et al., 2020). One of the most used biophysical parameter is the Leaf Area Index (LAI), which is derived from VIs. In its review, Salem et al. (2020) defines LAI as the area of one-sided leaf tissue per unit ground; this measures the density of the leaves surface in a canopy. Other often used biophysical parameters are Fractional vegetation cover (FCOVER) or fraction of photo-synthetically active radiation (FAPAR).

Finally, texture measures have also been proved to increase the prediction of AGB, in a few studies this happens to a greater extent compared to VIs (Pandit et al., 2019; Hall-Beyer, 2017*b*). Texture measures are based on a local variance function, on the image spatial resolution and on the size of the dominant objects, which helps identifying spectrally unique objects; the most often used in forest studies is the Grey-Level Co-occurrence Matrix (GLCM), which allows for the extraction of second-order texture measures (Haralick et al., 1973; Pandit et al., 2019).

Feature selection methods

AGB may have linear, non-linear or no relation at all with the model inputs, however, as a general rule, the features to be added in the regression model should have high correlation with AGB and low correlation among each others. Feature selection is often performed in order to rank which of the generated predictors would allow for a better performing regression model. Previous studies on AGB tend to follow two approaches: step-wise regression analysis or Random Forest derived measures (Gao et al., 2018; Lu, 2006).

Step-wise regression is often used to select features and it is simple to perform, however it is capable to account only for linear relationships (Gao et al., 2018; Lu et al., 2016). Random Forest (RF) is a Machine Learning (ML) non parametric algorithm, it is able to account for non-linear relationships between the features and AGB, and can provide a ranking of features based on their importance (Gao et al., 2018).

Regression models

Once the most adequate sensor and features are selected, the next important step is the choice of a regression model. Independently on the model, a satisfactory number of field measures are a prerequisite for the training and the validation of any supervised model (Lu, 2006).

In their survey on remote sensing-based AGB estimation methods in forest ecosystems, Lu et al. (2016) reviewed the most commonly used regression models in the field, starting with dividing them into parametric and non-parametric models. They found that parametric models are the most commonly used for estimating AGB, these are not data driven as they assume that the relationship between AGB and the selected predictors can be expressed a priori through parameters. As already mentioned in Section 2.3, the relationship between AGB and the predictors can be linear, non-linear or there can be no relationship at all; Parametric models which assume linear relationship are simple linear regression or multiple linear regression, and are often used for AGB estimation (Zheng et al., 2004; Wani et al., 2015; Nuthammachot et al., 2020), non-linear relationships can be accounted for by using a logistic regression model (Lu et al., 2016).

However, whenever an insufficient number of sample plots is available or the linear relationship between variables and biomass is weak, parametric model can result in a poor performance, which non-parametric models may lead to more accurate predictions (Lu, 2006). Non-parametric models are data-driven algorithms and do not make a priori assumptions on model structure (Lu et al., 2016). Commonly used non-parametric algorithms for AGB estimation are K-Nearest neighbor (kNN), RF, Support Vector Regression (SVR) and ANN (Coulibaly et al., 2008; Pandit et al., 2018; Gao et al., 2018; Puliti et al., 2020).

A few studies focus on the comparison among non-parametric and parametric models (Gao et al., 2018; Coulibaly et al., 2008). Gao et al. (2018) performed a comparative analysis between SVR, ANN, kNN, RF and Linear Regression (LR) using both a stratification and a non-stratification approach in a subtropical mountainous forest. The features selection was carried out with step-wise regression for the LR model, and with RF algorithm for the non-parametric models. ANN provided the best overall estimation in the non-stratification scenario with a RMSE of 27.6 t/ha, while LR provided a better estimation for broad-leaf and bamboo forests, when the stratification approach was used, with a RMSE of $24.9t/ha$ and $20.9t/ha$, respectively. Furthermore, this study found that LR is a valuable model with biomass ranging from 40 to 120 t/ha, however ANN led to better accuracy in biomass higher than 120 t/ha. Gao et al. (2018) concluded that there is not an unique modeling algorithm optimal for all forest types and AGB ranges, that is each scenario should be evaluated in its peculiarities.

Finally, once the model is trained, assessing its accuracy is the final step. Models are usually evaluated through the Root Mean Square Error (RMSE) and the Coefficient of determination (R2). However, the use of R2 has been proved by several authors to be an unreliable measure for evaluating non-linear problems, as explained in Appendix A. Furthermore, in a few studies other measures are included, such as relative RMSE, bias, standard deviation or p-value (Wani et al., 2015; Gao et al., 2018; Zheng et al., 2004;

Coulibaly et al., 2008; Vicharnakorn et al., 2014; Lu, 2006).

Chapter 3

Materials & Methodology

The following chapter contains a description of the study area, as well as data collection and preprocessing. Additionally, the methodology workflow is introduced. Each step of the methodology is covered in more detail in Chapter 4.

3.1 Study Area

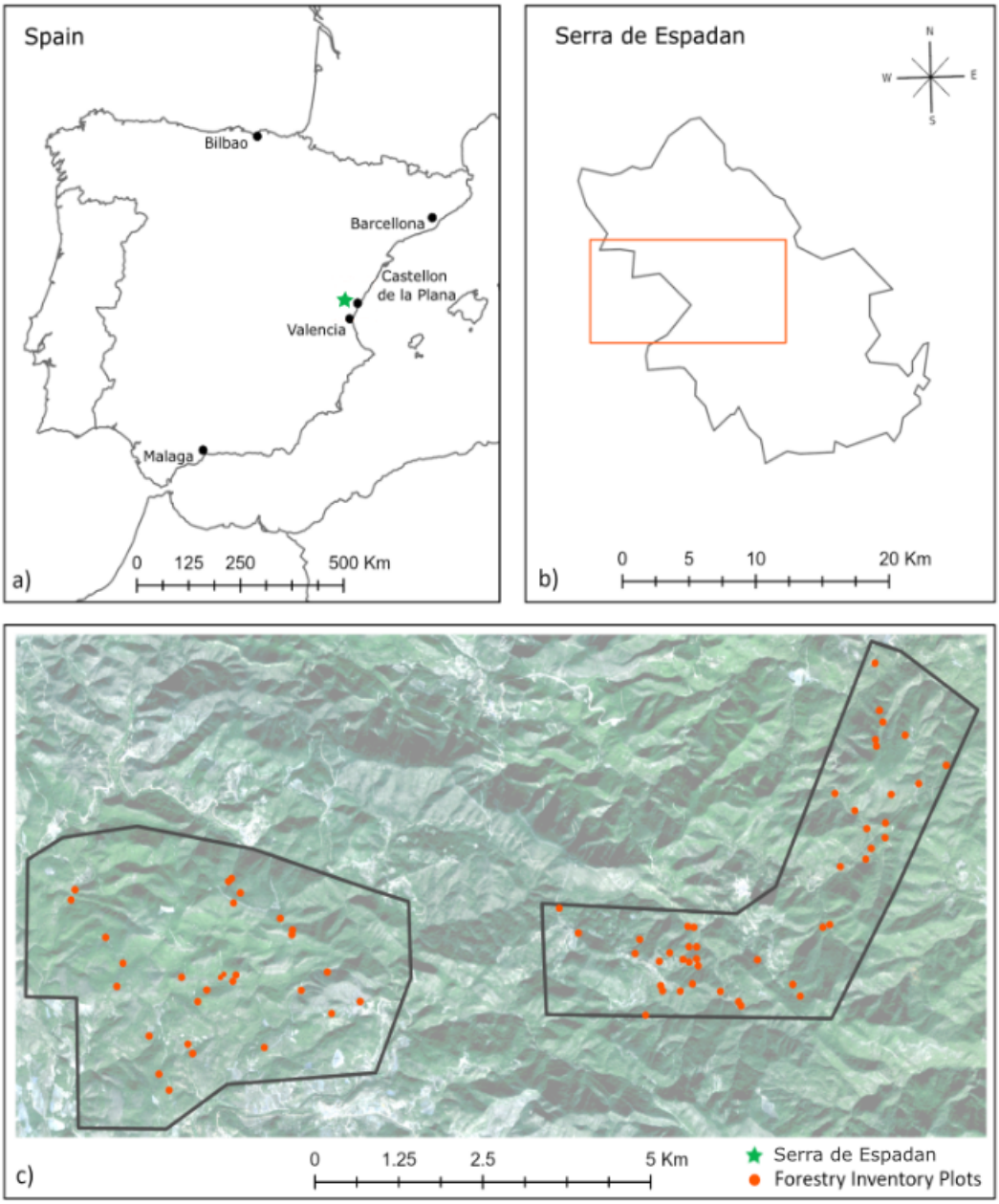
The following description of the study area was provided by a study from Torralba and Crespo-Peremarch (2018). The area object of study covers a total of 3741.5 ha and is located in the Natural Park of Serra de Espadan, in the eastern Spain province of Castellon (Figure 3.1). This natural park is a Mediterranean forest with soft and rounded hills, presence of abandoned farming with artificial terraces, and mountain peaks up to 1100 meters of altitude. An European Environment Agency report from 2017 classified this area as a semi-natural forest with a natural function, composition and structure, but modified by human activities throughout history (EEA, 2017). Forest types and conditions have been influenced by human needs and changes in land use, as well as reforestation of single species policies from the last century. This area displays a heterogeneous landscape dominated by pure and mixed native coniferous and deciduous forests, with species of the genera *Pinus* and *Quercus*.

In accordance to the global ecological zones described by FAO and to the Köppen-Trewartha Climatic groups, the Mediterranean climate is a variety of the subtropical climate, together with the Oceanic, the Humid subtropical, the Semi-desert, and desert climate (FAO, 2012). Hence, since many of the authors mentioned in Chapter 2 refer to their area of interest by using the climate domain - i.e. tropical, subtropical, temperate, boreal, polar - from now on, our area of study will be referred to by the mane of its major climate domain, that is subtropical.

3.2 Field data and field-based Above Ground Biomass

A field inventory with measured Above Ground Biomass (AGB) at the plot level was provided by the Geo-Environmental Cartography and Remote Sensing Group (CGAT) at the Universitat Politècnica de València (UPV) [1]; the collection of this forestry inventory was funded by the Spanish Ministerio de Economía y Competitividad, in the framework of the project CGL2016-80705-R. The field data was collected in September 2015 for a total of 73 circular plots with a radius of 15 m distributed throughout the study area. Diameter at Breast Height (DBH) and height were measured for trees with a DBH above 5 cm. For each species or forest type within a plot, AGB was estimated in $t/plots$ using species-specific and forest type-specific allometric equations from Gregorio et al. (2005). As showed in equation 3.1 and 3.2, the provided field-based AGB was then converted from $t/plots$ into t/ha.

[1] `http://cgat.webs.upv.es/`



Location of Serra de Espadan Natural Park, Spain (a). Location of the study area within Serra de Espadan Natural Park (b). Distribution of the forestry inventory plots within the area (c).

Figure 3.1: Study area and location of Forestry Inventory plots.

The field-based AGB ranges from a minimum of 0.35 *t/ha* to a maximum of 274.50 *t/ha*, with a mean value of 92.49 *t/ha*; the distribution appears right-skewed (Figure 3.2). It must be noticed that the values on the vertical axis in Figure 3.2 are referred to the status of data after preprocessing, that is once the forestry plots - now represented as polygons - are transformed into a feature point class where each plot results in 6 to 7 points; more information on this can be found in Section 3.5.

$$Area\ plot\ _{(ha)} = \frac{\pi r^2}{10000} \tag{3.1}$$

$$AGB\ _{(t/ha)} = \frac{AGB\ _{(t/plot)}}{Area\ plot\ _{(ha)}} \tag{3.2}$$

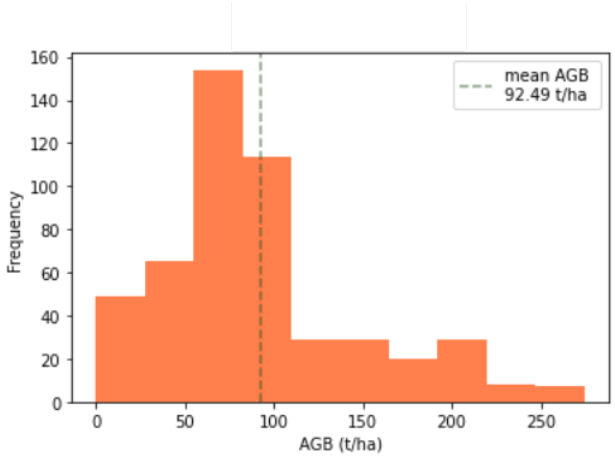


Figure 3.2: Distribution of the measured AGB.

3.3 Sentinel-2 collection and processing

The study area is covered by one single Sentinel-2A Level-1C tile. From the Copernicus Open Access Hub [2], 2 images were downloaded, so as to cover 2 seasons: Summer (August 2015 [3]) and Autumn (November 2016 [4]). Both dates are close to the time the forestry's inventory measures were collected and have a cloud coverage of less than 5%. However, clouds are absent in the spatial subset where the study area is located. This product is an ortho-image provided in the UTM/WGS84 projection, and composed of 13 spectral bands: visible, NIR, red-edge, and shortwave infrared (SWIR), which spatial resolution varies from 10, 20 and 60 m.

[2] `https://scihub.copernicus.eu/`
[3] S2A_MSIL1C_20150825T105046_N9999_R051_T30TYK_20200922T122045
[4] S2A_MSIL1C_20161107T105242_N9999_R051_T30TYK_20200923T092135

The first preprocessing steps were carried out using licence free tools [5] provided by the European Spatial Agency (ESA). Hence, in order to adjust the images from Top of the Atmosphere (ToA, L1C) to Bottom of the Atmosphere (BoA, L2A) reflectance, Sen2Cor was used; this plugin allows for atmospheric, terrain and cirrus correction. Further preprocessing was carried out using the Sentinel Application Platform (SNAP); red-edge and SWIR bands were re-sampled from 20 to 10 m using the nearest neighbor method; subsequently, the 3 bands with a spatial resolution of 60 m (band 1, 9 and 10) were excluded from the analysis by using the *subset* tool in SNAP, as these are mostly used for climate and atmospheric related studies. Finally, the remaining bands were clipped to the extent of the study area.

3.3.1 Vegetation Indices and biophysical parameters generation

Vegetation Indices (VI) and biophysical parameters have been proved to increase the performance of regression algorithms for AGB estimation in different ecological zones and forest types (Pandit et al., 2018; Forkuor et al., 2020; Galidaki et al., 2017). With the purpose of evaluating the performance of several feature selection methods, a wide range of VIs was generated using the calculator in SNAP. Particular care was taken in including VIs which required SWIR and red-edge bands, as shown in Table 3.1.

A total of 20 VIs were generated: Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1973), Green Normalized Difference Vegetation Index (GNDVI) (Gitelson et al., 1996), Soil Adjusted Vegetation Index (SAVI) (Huete, 1988), Modified Soil Adjusted Vegetation Index (MSAVI) (Huete, 1994), Global Environmental Monitoring Index (GEMI) (Pinty and Verstraete, 1992), Normalized Difference Vegetation Index red-edge 1, 2 and 3 (NDVIre1, NDVIre2 and NDVIre3) (Delegido et al., 2011), Chlorophyll red-edge index (Clre) (Gitelson et al., 2003) and Normalized Difference Water Index (NDWI) (Bo-cai, 1996). Hence, VIs were calculated for both dates, that is August 2015 and November 2016.

Furthermore, 5 biophysical parameters (Table 3.2) - Leaf Area Index (LAI), Canopy Water Content (LAI cwc), Canopy Chlorophyll Content (LAI cab), Fraction of absorbed photosynthetically active radiation (FAPAR) and Fraction of vegetation cover (FCOVER) - were calculated for each image by using the *biophysical processor* in SNAP. Such variables have been found to enhance the estimation of biomass by describing spatial distribution and dynamics of vegetation (Forkuor et al., 2020).

[5] `https://step.esa.int/main/toolboxes/`

Index	**Definition**	**Reference**
NDVI	(B8-B4)/(B8+B4)	(Rouse et al., 1973)
GNDVI	(B8-B3)/(B8+B3)	(Gitelson et al., 1996)
SAVI	[(1+L*)(B8-B4)]/(B8+B4+L*)	(Huete, 1988)
MSAVI	[B8+1-$\sqrt{(2B8+1)^2-8(B8-B4)}$]/2	(Huete, 1994)
GEMI	n**(1-0.25n**)-(B4-0.125)/(1-B4)	(Pinty and Verstraete, 1992)
NDVIre1	(B8-B5)/(B8+B5)	(Delegido et al., 2011)
NDVIre2	(B8-B6)/(B8+B6)	(Delegido et al., 2011)
NDVIre3	(B8-B7)/(B8+B7)	(Delegido et al., 2011)
Clre	(B7/B8)-1	(Gitelson et al., 2003)
NDWI	(B8-B12)/(B8+B12)	(Bo-cai, 1996)

L = 0.5, n**= (2(B8^2 − B4^2) + 1.5B8 + 0.5B4)/(B8 + B4 + 0.5)*

Table 3.1: Sentinel-2 generated Vegetation Indices

Biophysical parameter	**Definition**
LAI	Leaf Area Index
LAIcw	Canopy Water Content
LAIcb	Canopy Chlorophyll Content
FAPAR	Fraction of Absorbed Radiation
FCOVER	Fraction of vegetation Cover

Table 3.2: Sentinel-2 generated biophysical parameters

3.3.2 Texture measures generation

Once several spectral variables were generated - VIs and biophysical parameters -, further spatial predictors were included. This allows to consider not only the spectral response of different surfaces, but also the spatial relationships among these surfaces. For this purpose, texture measures derived from the Gray Level Co-occurrence Matrix (GLCM) were included, as they have been widely used for enhancing remote sensing-based classification and regression forestry-related problems (Pandit et al., 2019; Kelsey and Neff, 2014).

As Hall-Beyer (2017*a*) points out in her tutorial on GLCM texture, a GLCM is not an image, it is rather a tabulation expressing how often different combinations of Digital Numbers (Gray Levels) occur in an image band, showing all the possible combinations of value pairs and their frequency. This table is constructed by using each and every pixel of the image (reference pixel) and considering its neighbouring pixel or pixels (neighbour pixel). Texture measures derived from the GLCM consider the relationship between 2 pixels (GLCM) in the original image, as opposite to first order texture measures which are calculated directly from the original image pixels values without considering how this are related to one another (Haralick et al., 1973); most of GLCM derived measures used in Remote Sensing come from a series of papers of Haralick and colleagues in the 60s.

GLCM derived texture measures do not have a single way to be classified, hence, we are going to use the division used by Hall-Beyer in her tutorial (Hall-Beyer, 2017*a*); in this work, these measures are divided into 3 categories depending on the weights in the equations: measures related to contrast, measures related to orderliness and GLCM descriptive statistics. Pandit et al. (2019) explored the use of Sentinel-2 extracted GLCM texture measures in AGB estimation, and concluded that GLCM mean, variance, and dissimilarity from band 2 (blue), with a window size of 7x7, yield to the best AGB predictor for tropical and subtropical forests dominated by *Shorea robusta* and *Pinus roxburghii*, located in Parsa National Park, Nepal. In another study, AGB was estimated in the San Juan National Forest located in the southwest of Colorado; the area is characterised by *Ponderosa Pine* woodlands, Warm-Dry Mixed Conifer forests, Cool-Moist Mixed Conifer forests, and *Spruce-Fir* forests. In order to estimate AGB in such a context, a GLCM was derived from Landsat TM band 2; entropy, mean, and correlation were found to be the best predictors for that region (Kelsey and Neff, 2014). Thus, the importance of GLCM measures can vary based on sensor and study area, however, to minimize correlation among measures, Hall-Beyer suggests to use one measure from each category.

In the following study, 3 different GLCM texture measures were chosen, for a total of 12 new features to be tested.

These are divided as follows: contrast, entropy, and GLCM-mean were derived from both Sentinel-2 band-2 (blue) and NDVI generated for August 2015 and November 2016 to assess whether images from a certain season might provide better results (Table 3.3). The following definitions of contrast, entropy, and GLCM-mean used in remote sensing are taken from Hall-Beyer (2017*b*,*a*).

Entropy belongs to the category of measures related to orderliness, referring to how regular the differences between pixel value are within the window size - i.e. highly ordered images are characterized by frequently having the same combination of pixel pairs throughout

the image itself. For generating entropy texture measures, the weights in the equation are related to how many times a given pair occurs, by decreasing weights with the increasing of orderliness.

Texture measure	**Definition**	**Original band**
Entropy	$\sum_{i,j=0}^{N-1} iP_{i,j}(-lnP_{i,j})$	Band 2 NDVI
Contrast	$\sum_{i,j=0}^{N-1} iP_{i,j}(i-j)^2$	Band 2 NDVI
GLCM-Mean	$\sum_{i,j=0}^{N-1} iP_{i,j}$	Band 2 NDVI

P (i, j) is the normalized co-occurrence matrix such that SUM (i, j = 0, N-1) (P(i, j)) = 1 (Pandit et al., 2019)

Table 3.3: Sentinel-2 generated texture measures

Contrast belongs to the category of measures related to contrast, for which weights are dependent on the distance from the GLCM diagonal, with contrast increasing the further away from the diagonal. This results in pixel values of the generated measure increasing with the increasing of the contrast between two adjacent pixels in the original image.

Descriptive statistics derived from the GLCM work with the same principle as for common descriptive statistics, however in this case they are computed from the entries of the GLCM, whereas, common statistics are generated from the image original pixel values. That is GLCM-mean is not an average of the pixel values occurring in a window, but rather an average of how frequently certain pixel values combinations appear in the window.

Entropy, contrast and GLCM-mean were generated using SNAP software, setting a 7x7 window-size, 32 bit quantization level and a displacement of 1; this last measure indicates that the neighbour pixel is the one sited east of the reference pixel; higher displacement values mean that more than one neighbour pixel is considered in the calculations, however, previous studies proved that values higher than 1 do not lead to better performance (Pandit et al., 2019; Kelsey and Neff, 2014; Hall-Beyer, 2017*b*).

3.4 Digital Elevation Model collection

The European Digital Elevation Model (DEM) and derived slope were downloaded in the section *Imagery and Reference Data* of the Copernicus website [6]. These products have a spatial resolution of 25m. DEM and slope were re-sampled to the same spatial resolution as the Sentinel-2 images (10m) by making sure that cell size and cell positioning matched. DEM up-sampling is carried out in ArcGIS Pro using the *nearest neighbor assignment* method, since it does not alter the input cell value. Furthermore, DEM and slope are clipped to the study area extent.

3.5 Methodology workflow

Once every feature - VIs, biophysical parameters, texture measures, spectral bands, DEM and slope - was generated or collected, re-sampled to 10 m, and clipped to the study area extent, a total of 63 features was transformed from their original raster format to vector format (points) by using the *Raster to Point (Conversion)* tool in ArcGIS Pro; subsequently, the *Extract multiple values to points* tool was used in order to create a final feature point with associated table containing a total of 64 columns, representing the 63 features and the geometry field. This last column allows for the feature point to be transformed back into a raster format so as to subsequently generate the final maps.

In parallel, the circular forestry plots were converted from the vector format (polygons) into raster cells, as represented in Figure 3.3, a) and b). Pixels with the measured AGB as value, were generated only when more than 50% of the area of the reference pixel was covered by the forestry plot; a Sentinel-2 10 m preprocessed image was used as reference pixel. Additionally, this raster layer was transformed back into a vector (points) as shown in Figure 3.3, c).

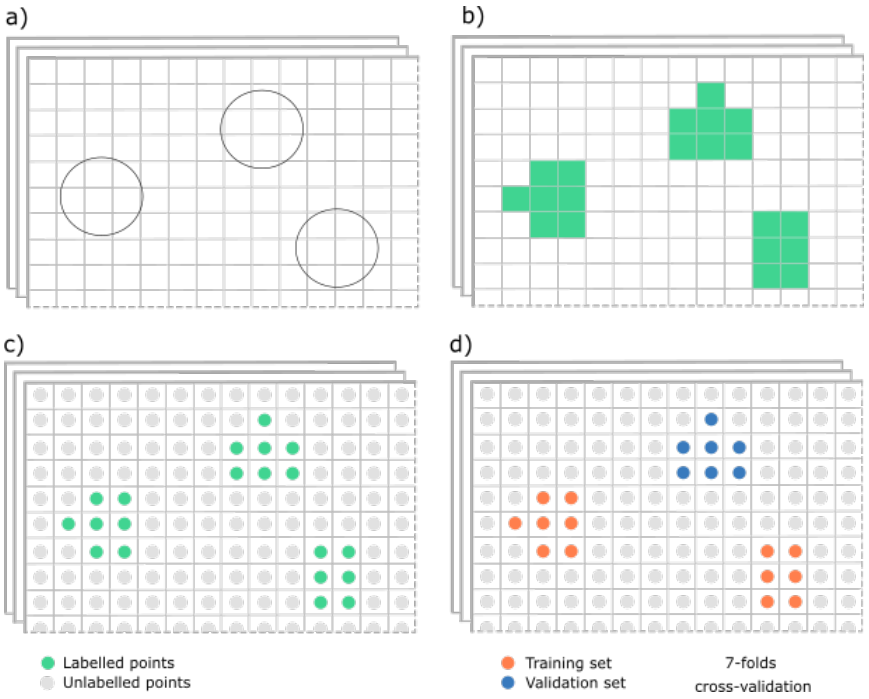
Henceforth, all the analysis are performed using the generated vector file (points) composed of 65 columns, of which 63 represent each of the generated features, 1 represents the field-based AGB values, and the last one represents the geometry field.

Before proceeding with the analysis, the 63 features were normalized. This step is necessary as the features differ in both range and unit of measure. Moreover, certain regression models as well as feature selection methods can result in a sub-optimal performance when the inputs are not normalized, in the sense that such algorithms tend to give more importance to high cardinality and continuous features.

The 63 features were normalized using the RobustScaler from Scikit-learn [7]. The RobustScaler normalizes the dataset according to the interquartile range, such an approach offers a good handle of outliers compared to mean- and variance-based normalization methods (Pedregosa et al., 2011).

[6] `https://land.copernicus.eu/`

[7] `https://scikit-learn.org/stable/index.html`



Circular forestry plots in vector format (polygon) (a). Rasterization of the forestry plots (b), vectorization to points of each layer including forestry plots (c), and division in training and validation set (d).

Figure 3.3: Explanation of preprocessing steps.

Dividing the labelled portion of the dataset into training and validation set using automatic methods could lead to overoptimistic results; with this approach labelled points belonging to the same forestry plot can be automatically parted to be part of both the training and the validation set, this issue is also known as data pollution. With the purpose of generating reliable results, the author built a function which performs data normalization with the RobustScaler and a 7-folds cross-validation by posing 2 constraints within the function. Firstly, points belonging to the same original forestry plot can be included exclusively in the training or in the validation set (Figure 3.3, c) and d)). Secondly, only the training set is used to fit the scaler, this allows validation data to remain unseen throughout the process. The entire dataset is finally normalized utilizing the trained scaler. This function is used before each step, that is feature selection, feature selection testing, and hyper-parameters fine-tuning.

As briefly mentioned in Chapter 1 and illustrated in Figure 3.4, the 63 features are going to be ranked using 4 different methods - Mean Decrease in Accuracy (MDA), Mean Decrease in Impurity (MDI), L1 Regularization (LASSO) and Principal Component Analysis (PCA). The features ranking provided by these methods is then going to be tested by implementing different regression algorithms: Random Forest (RF), K-Nearest Neighbour (KNN), Extreme Gradient Boosting (XGB), and 3 Artificial Neural Networks (ANN) characterized by various levels of complexity. Hyper-parameters fine-tuning is then performed with the purpose of enhancing the performance of KNN, RF and XGB. Subsequently, the final number and type of features, as well as the final predictive model with opti-

mal hyper-parameters are going to be chosen based on the model performance, expressed in Root Mean Squared Error (RMSE). The RMSE is calculated using the equation 3.3, where y_p is the predicted AGB of a n_i point, y_o is the observed AGB of the n_i point, and n is the number of validation points. Besides, it must be mentioned that the coefficient of determination (R^2) will not be used for model evaluation since it has been proved to be an unreliable measure for evaluating non-linear problems; more information are found in Appendix A.

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}\Big(y_p - y_o\Big)^2} \tag{3.3}$$

Finally, the best performing model and features are going be used for generating the final maps, as well as for performing model explanation with SHapley Additive exPlanations (SHAP), recently proposed by Lundberg et al. (2020) for tree-based Machine Learning models.

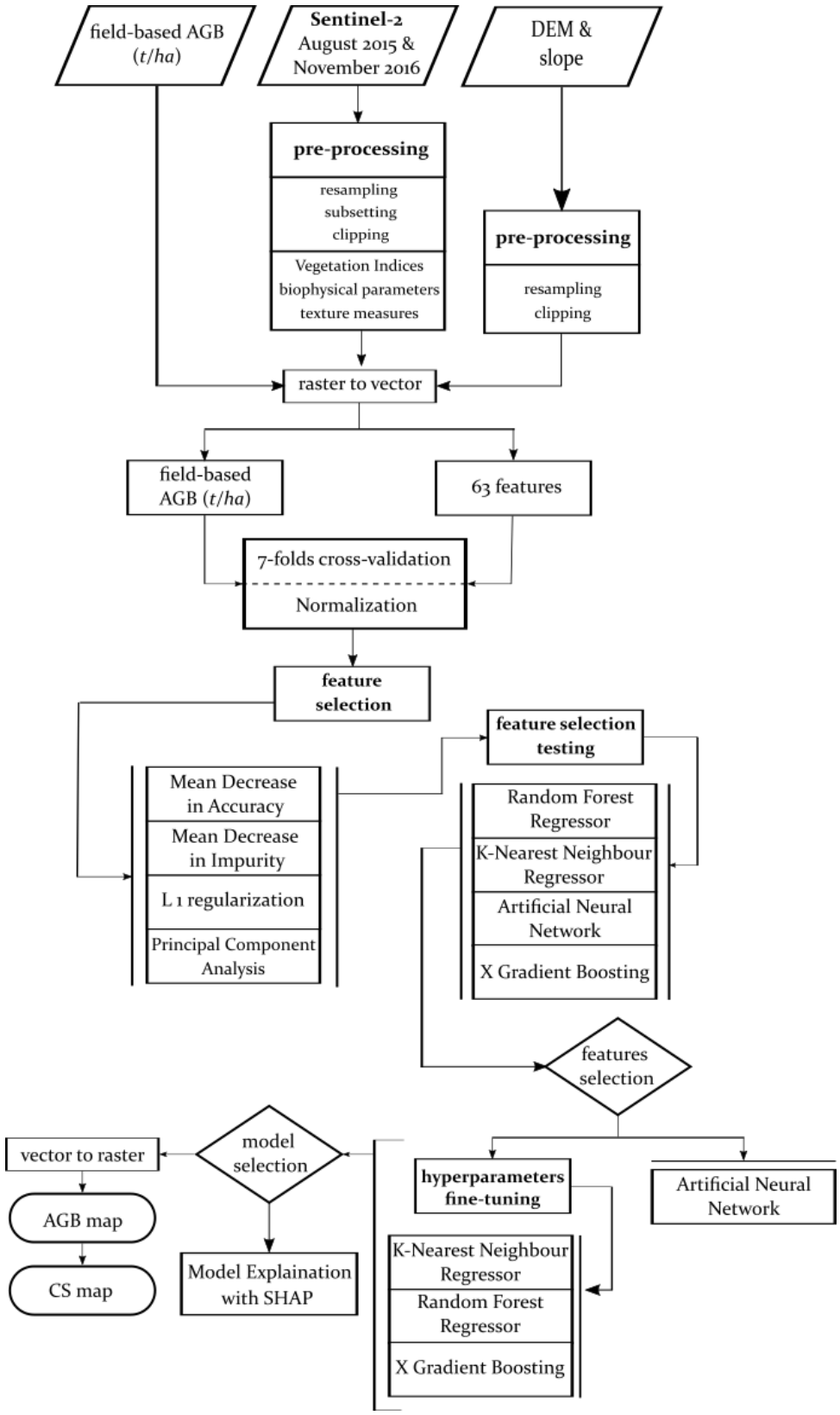


Figure 3.4: Methodology workflow.

Chapter 4

Implementation & Results

The following chapter covers the implementation and results of this study. Specifically, we are going to answer each of the Research Questions (RQ) stated in Chapter 1. Both the preprocessing steps and following analysis in this Chapter were performed by scripting in Python3, a licence-free general-purpose programming language. Additionally, the final maps were generated using ArcGIS Pro software, version 2.15.

4.1 Features selection

In order to build a reliable model for the estimation of Above Ground Biomass (AGB), careful evaluation is required when deciding which and how many features to include in the predictive model. To the best of our knowledge, step-wise regression and Random Forest (RF) measures are amongst the most frequently used approaches for feature selection in the field of AGB (Zheng et al., 2004; Vicharnakorn et al., 2014; Puliti et al., 2020; Forkuor et al., 2020).

Despite step-wise regression is nowadays still widely used for this purpose, the author decided not to include it. Such a decision is legitimized by the work of several authors proving its unsuitability for feature selection (Babyak, 2004; Smith, 2018). By going through a series of automatic steps, step-wise regression chooses a number of features from a group of available potential variables; this is done via either forward, backward or bi-directional elimination; at each step, features are evaluated typically through a t-statistic test (Efroymson, 1960) and either eliminated or included in the model. Authors from different fields affirm its unsuitability for feature selection, pointing out the misapplication of what is supposed to be a single-step statistical tests to a multi-step procedure, this was proved to lead to over-fitting and to create a false confidence in the final model. Smith (2018) carried out an experiment by introducing false variable into an existing data-set; his work shows how step-wise regression may choose nuisance variables rather than true variables and that the out-of-sample accuracy of the model may be far worse than the training accuracy. Smith (2018) concludes that such issues exacerbate when a larger number of features is tested.

A total of 4 features selection methods are evaluated, including both supervised and unsupervised approaches. Other than the Mean Decrease in Impurity (MDI) and the

Mean Decrease in Accuracy (MDA), this work explores the potential of L1 Regularization (LASSO); both of which are supervised methods - i.e. the target is known. Hence, an unsupervised method, Principal Component Analysis (PCA) is included, and its performance evaluated.

4.1.1 Random Forest for feature selection

RF is an ensemble Machine Learning (ML) algorithm constituted of several decision trees (Breiman, 2001), and it is widely used for classification, regression, and other tasks. This algorithm is often used for feature selection as it contains built-in methods able to generate a ranking of the best features to include in the model. In this work, feature selection with RF is performed by using 2 of these methods: Mean Decrease in Impurity (MDI) and Mean Decrease in Accuracy (MDA). Most authors either use the MDI, an impurity based measure or MDA, a permutation-based measure (Pandit et al., 2018; Puliti et al., 2020; Hernandez-Stefanoni et al., 2020).

A RF was modelled by using all the 63 features, the measured AGB, and default hyperparameters showed in Table 4.5. Hence, the dataset was normalized and cross-validated by using the previously described cross-validation function with 7 folds (Section 3.5).

Mean Decrease in Impurity (MDI)

The RF algorithm provides 3 measures of impurity: Gini index, entropy, and variance. Variance, or residual sum of squares, is used to measure node impurity in regression problems, and it represents the total reduction of the variance of the target variable due to the split of a certain feature at the node (Lewinson, 2019). Impurity measures can be used for feature selection by evaluating the extent to which each feature contributes to decreasing the averaged impurity in each tree composing the forest (Lewinson, 2019), so as to calculate the MDI for each feature. The feature able to account for more variance decrease is going to be at the top of the ranking (Lewinson, 2019). Therefore, the MDI can be see as the total decrease in node impurity from splitting on the variable, averaged over all trees (Hong Han et al., 2016).

Despite impurity-based importance evaluation being widely used for feature selection, this has been proved to be biased towards continuous or high cardinality features (Hong Han et al., 2016). Another often mentioned critique about this approach is that impurity is not a direct measure of a model performance, and it might lead to a false confidence in the chosen variables (Lewinson, 2019).

Figure 4.1 shows the MDI that each feature brings to the model. We can observe that the first 9 selected features, apart from the DEM, were all extracted from the Summer image. Furthermore, biophysical parameters were often selected, specifically Canopy Chlorophyll Content (LAIcb), Canopy Water Content (LAIcw), and a chlorophyll index calculated using red-edge bands (Clre). A SWIR band (b12), the DEM, and a texture measure, the entropy measure derived from the summer NDVI, were also included in the top selected

features.

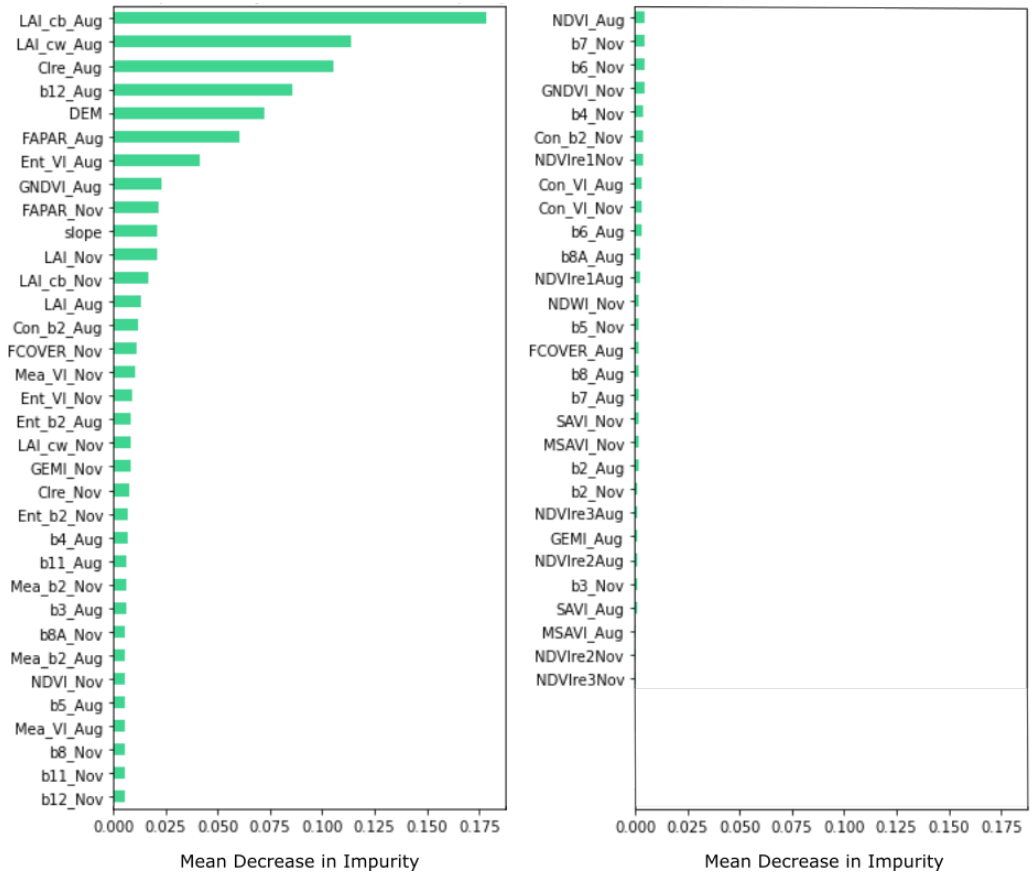


Figure 4.1: Feature selection based on the MDI measure.

Mean Decrease in Accuracy (MDA)

Permutation-based feature importance, henceforth MDA, performs a ranking of features based on the increasing mean error of a tree in the forest when the feature values are randomly shuffled though maintaining their distribution (Genuer et al., 2010). For regression problems, the error refers to the Mean Squared Error (MSE). MDA is performed by measuring the impact of each feature on the model MSE by starting from a MSE baseline (Altmann et al., 2010). On the downside, permutation-based measures tend to overestimate the importance of highly correlated features (Lewinson, 2019).

The general idea is to train a baseline model with all the original features and computing its MSE with unseen data; the values of one of the features are then permuted and the change in error is measured by subtracting the baseline MSE with the MSE generated after the feature permutation. This is repeated for every feature. When a non-important feature is permuted, this has little to none influence on the error, while when an important feature is permuted, this corresponds to an increases of the MSE. Features which permutation increases the validation error the most, will be found at the top places of the features ranking.

Figure 4.2 illustrates the feature ranking performed using MDA. Looking at the first 10

features it can be noticed how the order is slightly different than the one individualized with the MDI method. The model performance suffers the most when biophysical parameters from the Summer image (LAI cb, LAI cw and Clre), the DEM and the SWIR band (b12) are permuted. The Entropy measure (Ent VI Aug) and the fraction of absorbed photo-synthetically active radiation (FAPAR Aug and Nov) were also included among the most influential features, as it happened in the MDI ranking. Overall, permutation feature importance is privileging features extracted from the Summer image.

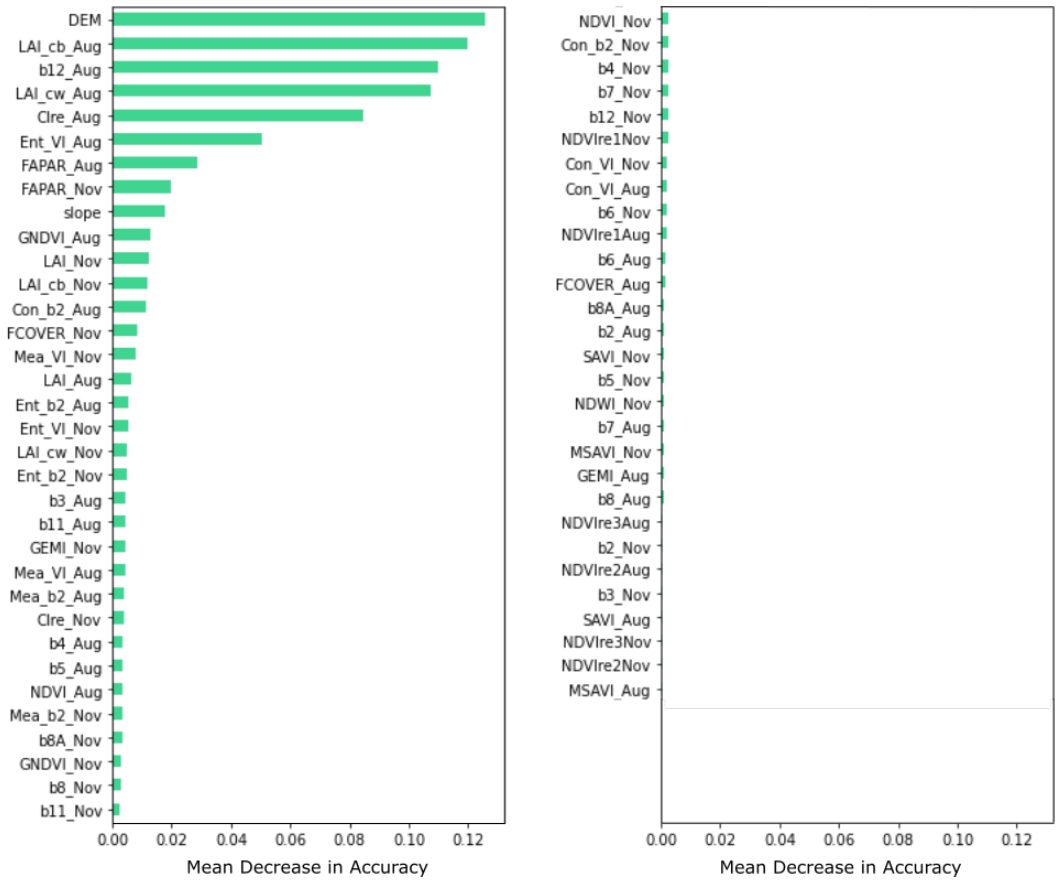


Figure 4.2: Feature selection based on the MDA measure.

4.1.2 L1 Regularization for feature selection

Regularization algorithms aim to minimize the residual sum of squares of the model by using the Ordinary Least Square (OLS); this is achieved by shrinking the estimated regression coefficients approaching to zero (Muthukrishnan and Rohini, 2016). The Least Absolute Shrinkage and Selection Operator (LASSO) was proposed by Tibshirani (1996); LASSO regression is an L1 Regularization technique, which minimizes the absolute sum of the coefficients; while L2 regularization minimizes the squared sum of the coefficients (Muthukrishnan and Rohini, 2016).

Feature selection with LASSO is performed by multiplying every coefficient by a constant, that is the regularization parameter *alpha* (α). *Alpha* controls the strength of penalty, meaning that for large values of *alpha*, a large number of coefficients is forced to zero,

which corresponds to the exclusion of those features from the model; on the other hand, when *alpha* is equal to zero, we have the classic OLS (Fonti, 2017).

LASSO has been proved to perform well in case of multicollinearity, that is if there is high correlation between the features, LASSO chooses only one of them and forces the others to a coefficient of zero (Muthukrishnan and Rohini, 2016). In other words, regularization consists in adding penalty to features introduced in the model in order to decrease the freedom of the model and avoid over-fitting. In linear models this penalty is added in the weights (or coefficients) that multiply each features. Working with LASSO gives us the ability of bringing weights all the way down to zero, so as to drop less relevant features from the model, while those which weights are not zero are added into it.

By keeping in mind that increasing the penalty *alpha* (α) will increase the number of features removed, we need to keep an eye on this parameter and monitor that we do not set a penalty too high so that it removes important features, or too low so that does not remove unimportant features.

Therefore, in order to better explain LASSO and its regularization parameter, it helps having a look at the classic linear model. This searches for weights to associate to each feature $(w_0, w_1, w_2 \dots w_n)$, such that:

$$X_0 w_0 + X_1 w_1 + X_2 \ddot{w}_2 + \ldots + X_n w_n = y \tag{4.1}$$

where y is the target and X_i are the input features. Hence, the error in the linear model is given by:

$$error = \|y - X_0 w_0 + X_1 \ddot{w}_1 + X_2 w_2 + \ldots + X_n w_n\| \tag{4.2}$$

LASSO introduces a regularization term, which penalizes the model for high value of the weights.

$$error = \|y - X_0 w_0 + X_1 w_1 + X_2 \ddot{w}_2 + \ldots + X_n w_n\| + \underbrace{\alpha \sum |w_i|}_{regularization\ term} \tag{4.3}$$

The regularization term can be seen as the opportunity cost of choosing to use a given feature in the model, and the regularization parameter, α, controls how much we value this cost. Therefore, for high values of α the model will select a small number of features

because the opportunity cost of including each of them increases. On the other hand, by relaxing the regularization term - i.e. lowering the value of α - the model starts including more and more features in an order of importance.

This behaviour can be better understood by observing Figure 4.3, showing how the number of features included in the model decreases as α increases from 0 to 10.

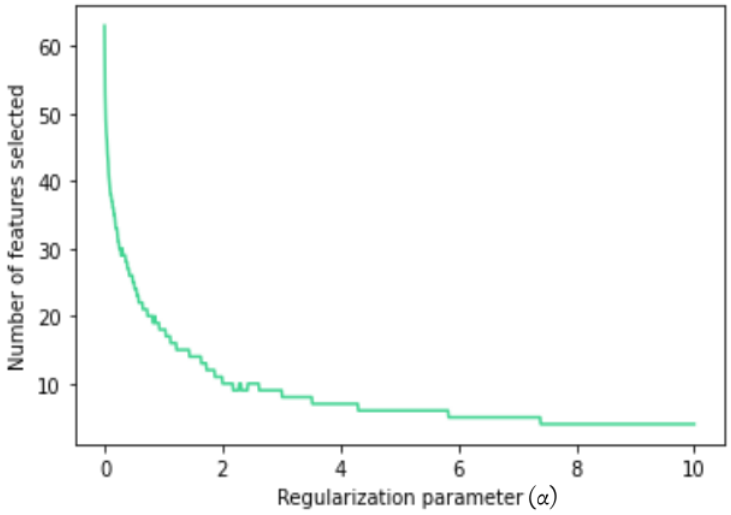
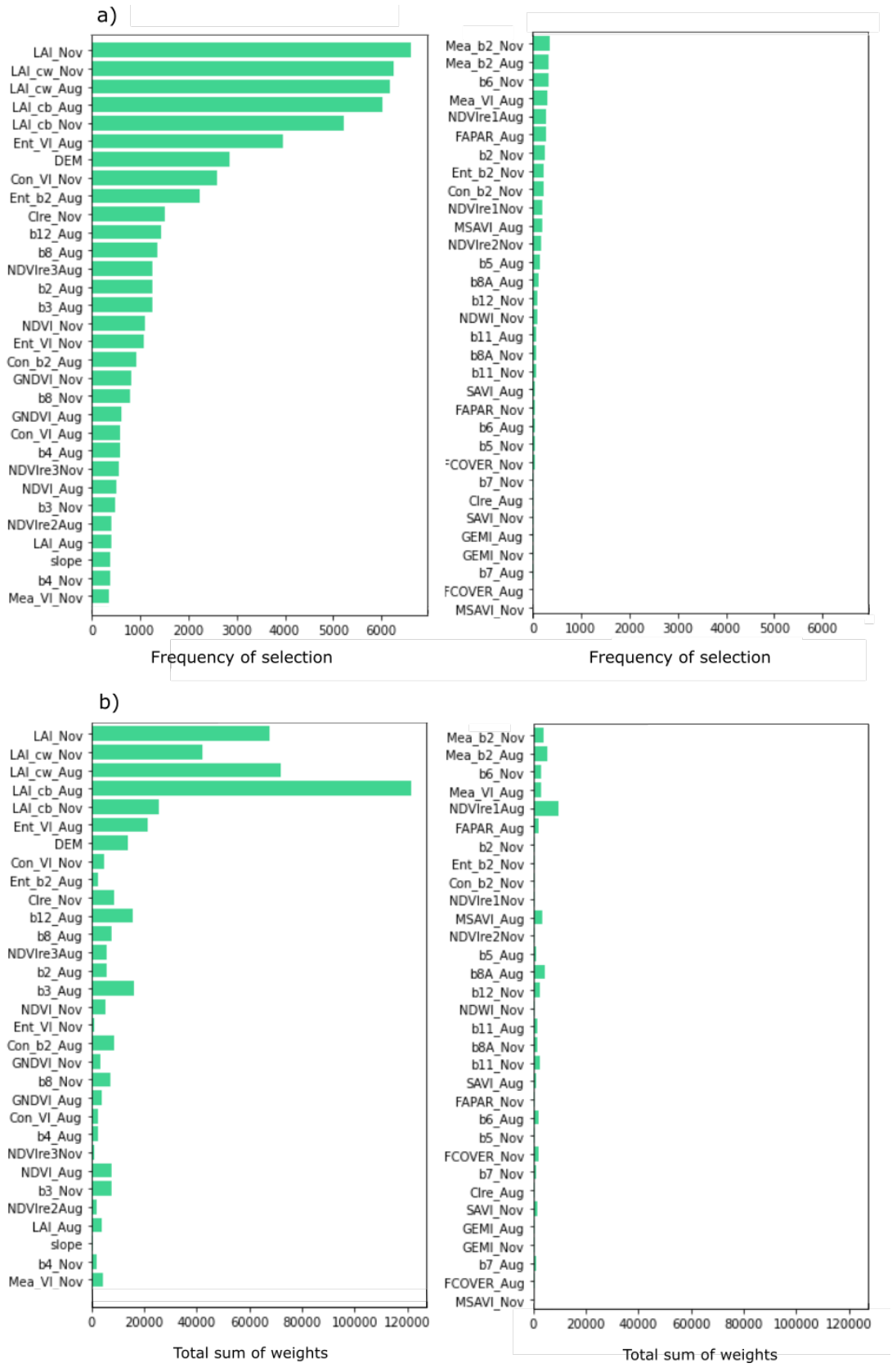


Figure 4.3: Regularization parameter and number of selected features.

In this work, LASSO was computed 1000 times, by ranging the value of α from 0 to 10, with 1000 intervals. A ranking of the selected features was finally generated by doing a sum of the frequency of their appearance throughout the 1000 iterations, as shown in Figure 4.4, a). For instance, the first 3 selected features (LAIcwAug, LAIcwNov and LAIcbAug) were included in the model for every value of α, while the forth one (LAINov) was included 950 times out of the 1000 iterations. Moreover, Figure 4.4, b), shows the sum of each feature coefficient (or weight) throughout the 1000 iterations. Overall, the sum of weights tends to be higher for the most frequently included features.

Finally, LASSO frequently selected biophysical parameters from both August 2015 and November 2016, as well as texture measures (Entropy and Contrast) and the DEM.



Frequency of selection of each feature throughout the 1000 iterations (a) and sum of each feature's weight throughout the 1000 iterations (b); features are ordered based on their frequency of selection.

Figure 4.4: Feature selection based on LASSO.

4.1.3 Principal Component Analysis for feature selection

It is well-known that feature selection is not the primarily use of Principal Component Analysis (PCA) (Walker, 2019). However, the author decided to incorporate it in this work so as to explore the performance of an unsupervised approach.

PCA performs dimensionality reduction by transforming a coordinate system into another in a way to capture the variation in the data, that is preserving as much information present in the original dataset as possible (Guo et al., 2002). After its implementation, the first dimension, henceforward component, in the new coordinate system is a linear combination of the features and it has the maximum variance, while the second dimension represents the maximum of the remaining variance and so on; in dataset composed of a small number of features, the first few components usually explain most of the dataset variance (Malhi and Gao, 2004).

It is considered acceptable to use PCA as a feature selection method if features are ranked based on the amount of variance they explain in each component (Walker, 2019). That was here achieved by using the loadings, which can be seen as the weight that each feature has in a component of the PCA. Variables with higher loading values have higher influence in explaining the variance of the dataset.

Visually, this can be observed in Figure 4.5, which displays the loading values of each feature within the first Principal Component. However, the first Principal Component only accounts for 50% of the total variance of the dataset; while first and second Principal Components together add up to explaining circa 65% of it (Figure 4.6). With the purpose of finding a sufficient number of Components, the percentage of variance explained by adding an increasing number of them was tested. As showed in Figure 4.6, the total explained variance eventually starts to converge as we add more components; the first 10 Principal Components already explain 95% of the total variance. Moreover, it can be noticed in Figure 4.7 that after reaching 10 Principal Components the feature loadings and ranking tend to stabilize when further Components are included; 10 was then chosen as an appropriate number of Components for unsupervised feature selection.

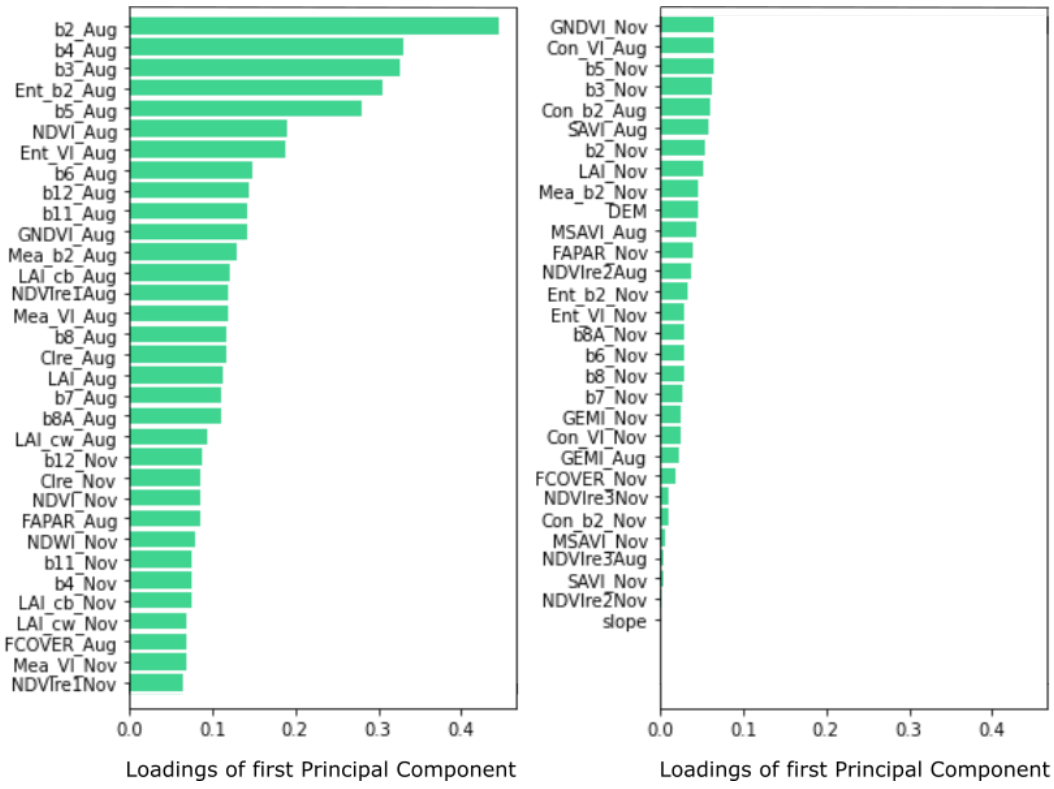


Figure 4.5: Loading values for the first Principal Component.

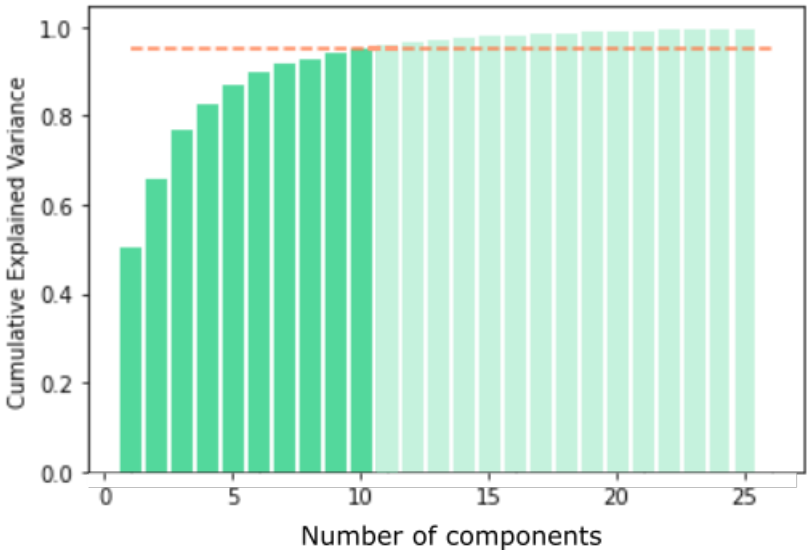


Figure 4.6: Cumulative explained variance.

Following this logic, the feature ranking was generated by doing a weighted sum of the feature loadings over the 10 Principal Components, weighting them by the explained variance ratio of each Component.

In Figure 4.7 a), we can observe that the features with higher loading values are Entropy measures from August 2015 extracted from both band 2 (Ent b2 Aug) and the Normalized

Difference Vegetation Index (Ent VI Aug), as well as band 2, band 3, band 4 and band 5 from the month of August (b2 Aug, b3 Aug, b4 Aug, b5 Aug). Features extracted from the month of November as well as topographic measures are lower in the ranking.

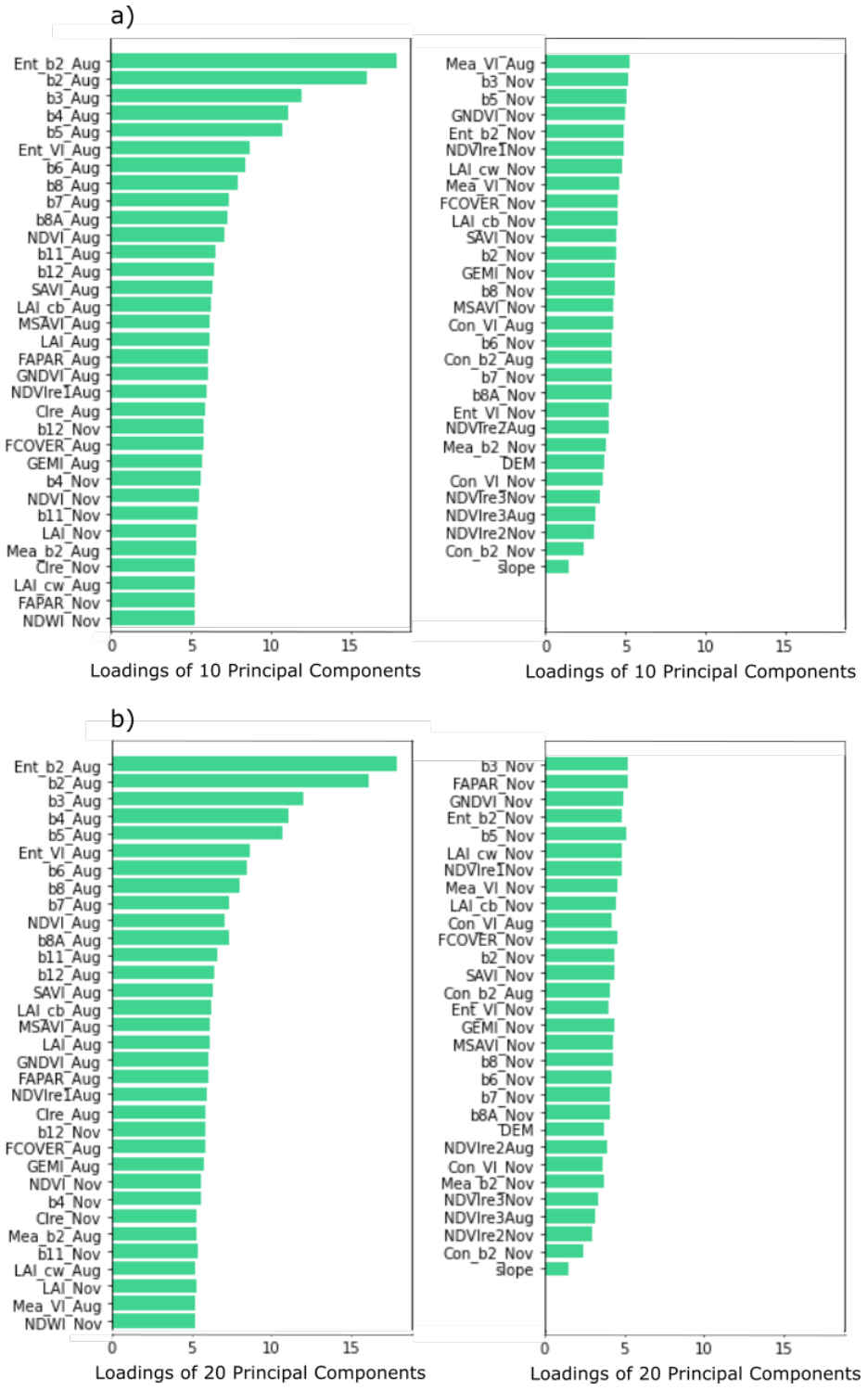


Figure 4.7: Loading values of features using 10 Principal Components (a) and Loading values of features using 20 Principal Components (b).

4.2 Feature selection testing

In this section, the previously explained feature selection methods are tested and evaluated using several non-parametric Machine Learning (ML) algorithms: K-Nearest Neighbour (kNN), Random Forest (RF), Extreme Gradient Boosting (XGB) and, lastly, 3 Artificial Neural Networks (ANN).

4.2.1 Feature selection testing with K-Nearest Neighbour

kNN is a ML algorithm and it does not assume normal distribution or linear relationships; furthermore, it can be used for regression and classification problems. This non-parametric algorithm has been widely tested for the prediction of AGB (Gao et al., 2018; Salem et al., 2020). With "k" representing the number of neighbours taken amongst the points belonging to the training set. In the following section the default value of 5 was used for "k", as indicated in Table 4.4. The estimation of the queried points is given by the weighted average of the k nearest neighbours. The weights are set higher for those neighbours spectrally similar to the queried point; while, smaller weights are set for those neighbours spectrally less similar to it.

In order to test the 4 features rankings, several kNN regression models were trained. More specifically, one kNN model was generated for each combination. For instance, the performance of the features ranking resulted from the MDI measure (Figure 4.8) was evaluated by testing 63 models with the first one having as input only the first-place ranked feature (LAI cb Aug), the second model having as inputs only the first two top-ranked features (LAI cb Aug and LAI cwAug), and so on, until the last model which was built using as inputs all the 63 features. Hence, each model was cross-validated by using the cross-validation and scaling function defined in Chapter 3.

The performance of the implemented models is illustrated in Figure 4.8, a); this shows how the RMSE behaves when an increasing number of features is included in the model in a certain order. Overall, it can been observed that, after an initial arise of the performance i.e. decrease of the RMSE, the error started steadily increasing once more than circa 25 features are included in the model; the error converges when all the 63 features are included, independently on the feature selection method. Furthermore, PCA was, at every point, the worst performing method, while feature selection with MDA, MDI and LASSO brought to similar errors. The lowest RMSE (43.93 t/ha) was reached when the first 25 features selected by MDA were included; these features are summarized in Table 4.4.

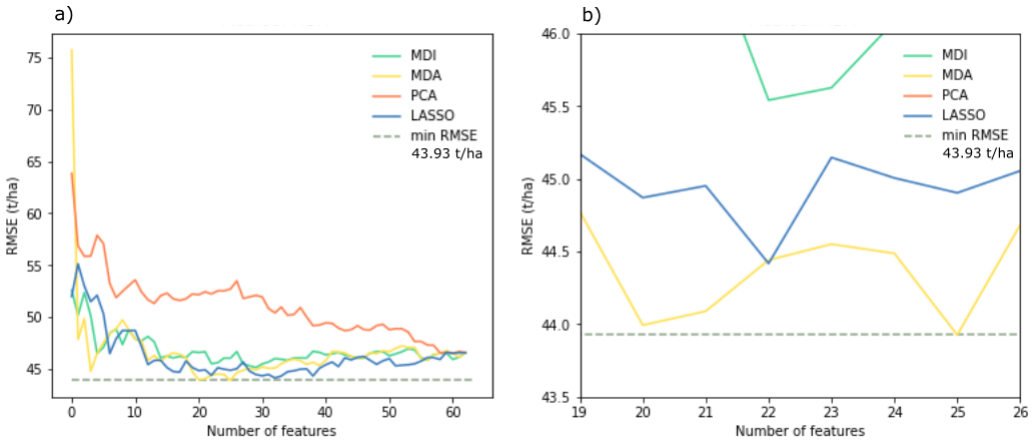


Figure 4.8: Feature selection methods testing with K-NN (a) and magnification over the best performing number of features (b).

In Figure 4.9, the measured AGB values of each validation folds were plotted together with the model predictions. It can be noticed that low and high values were over- and underestimated, respectively.

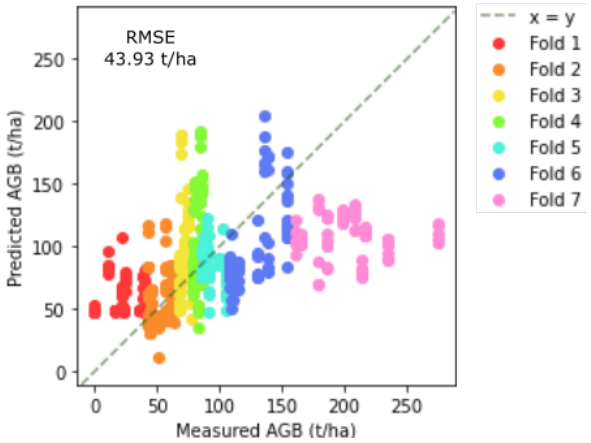


Figure 4.9: Goodness of fit on unseen data with kNN.

4.2.2 Feature selection testing with Random Forest

In order to test the 4 feature selection rankings, several RF algorithms were trained using default hyper-parameters (Table 4.5). One RF was generated for each features combination with the same logic as for the kNN, explained in Section 4.2.1. Hence, each model was cross-validated by using the cross-validation and scaling function defined in Chapter 3.

The performance of the implemented RFs is illustrated in Figure 4.10, a); this shows how the RMSE behaves when an increasing number of features is included in the model in a certain order. Overall, it can been observed that, after an initial arise of the performance

i.e. decrease of the RMSE, the error started steadily increasing once more than circa 5 features were included in the model.

When analysing the same graph in more detail, it can be observed that the RF modelled using the features selected by PCA hold the highest RMSE independently on how many of them were included in the model. Furthermore, the magnified plot in Figure 4.10 b), shows that the lowest RMSE corresponds to 44.29 t/ha and it is achieved by using the first 3 features of the ranking generated by MDA.

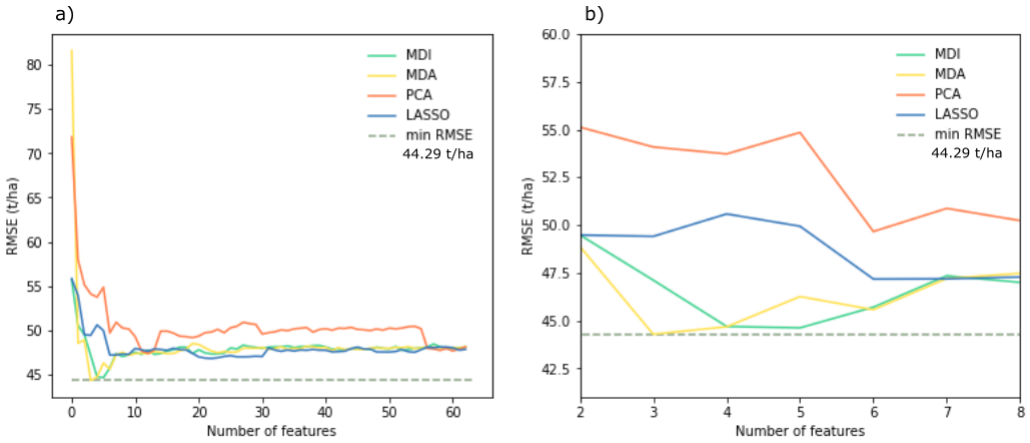


Figure 4.10: Feature selection methods testing with RF (a) and magnification over the best performing number of features (b).

Finally, the measured and unseen AGB values were plotted together with the prediction for each folder of the RF model yielding to the lowest RMSE amongst those tested (Figure 4.11). Here, it can be noticed that values lower than 40 t/ha and higher than 160 t/ha were over- and underestimated, respectively.

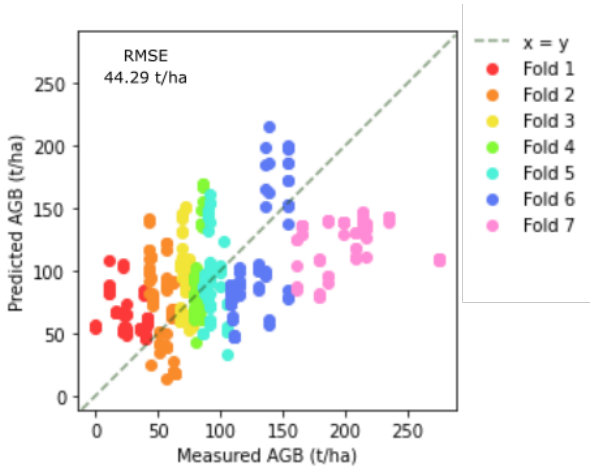


Figure 4.11: Goodness of fit on unseen data with RF.

4.2.3 Feature selection testing with Extreme Gradient Boosting

Increasing the predictive power of a regression model, by decreasing its validation error, is often achieved through performing either Regularization, Boosting or Bagging techniques (Sutton, 2005).

In Random Forest (RF), bagging is used to average the results of the trees composing the forest. Bagging involves the generation of a bootstrap dataset by randomly selecting samples from the original dataset, and averaging the prediction of the single trees (Breiman, 2001); in the RF algorithms each tree is independent of the others and has the same weight on the final model predictions. On the other hand, in the Boosting algorithm, single trees depend on each other. Boosting was first introduced by Schapire (1990), and it works by transforming several so called weak learners in a strong learner so as to improve the model performance; this is achieved by iteratively adding individual weak learners, in this case decision trees, into a stronger model. Weak learners can be defined models which output is only moderately correlated with the true value, while still yielding a better performance than using random guesses (Schapire, 1990). These trees are not built on random subsets of the original dataset as for the RF algorithm, but rather each new tree is built in a way to focus on those instances affected by bigger errors, so as to take into account the previous tree mistakes. This is possible by adjusting the weights that each instance has, starting with the same weight all-over the dataset and increasing it for those instances affected by high prediction errors.

Gradient Boosting and Adaptive Boosting are two often used Boosting algorithms. In this work, the focus is on an implementation of Gradient Boosting, namely Extreme Gradient Boosting (XGB), recently introduced by Chen and Guestrin (2016). The Gradient Boosting algorithm combines the Boosting principles with the gradient descent. Each time a new tree is built, the gradient is utilized for minimizing a loss function; the predictions of the tree are compared with the correct output, and the errors are used for calculating the gradient. This describes the steepness of the error function and it is found by doing the partial derivative of the loss function (Friedman, 2002), such measure is utilized for boost the model predictions by adjusting the prediction of the next tree, and so on (Friedman, 2002); the learning rate can be seen as the speed at which the gradient descends.

The Extreme Gradient Boosting (XGB) is amongst the most popular variants of the Gradient Boosting algorithm and it has been the winner of many Machine Learnig (ML) competitions (Seif, 2019); Gradient Boosting can be seen as one of the parts composing this algorithm. XGB is a big ML algorithm composed of many parts such as gradient boosting, regularization, cache aware access, variables splitting into quantiles, and so on. Here, only 3 of its major differences with the classic Gradient Boosting are mentioned. Firstly, Regularization using Ridge (L2) and LASSO (L1) regression is performed each time a new tree is added, this decreases the importance given to the training set so as to improve the model generalization which allows a better performance on unseen data (Chen and Guestrin, 2016). Secondly, instead of computing the first partial derivative, XGB speed is increased by computing the second partial derivative, or second order gradient, of the loss function (Chen and Guestrin, 2016). Lastly, XGB allows for quick computation timing; this is possible thanks to a cache aware access which allows to utilize cache memory for calculating similar scenario without the need of accessing the slower

hard drive memory,

In order to test the 4 feature selection rankings, several Extreme Gradient Boosting algorithms were trained using default hyper-parameters, as showed in Table 4.6. Hence, each model was cross-validated by using the cross-validation and scaling function defined in Chapter 3.

The performance of XGB is illustrated in Figure 4.12, a). Overall, it can been observed that, after an initial arise of the model performance, the error started erratically increasing once more than 23 features were included in the model. Moreover, the XGB algorithm modelled using the features selected by PCA held the highest RMSE independently on how many features were included in the model. Hence, the magnified plot in Figure 4.12 b), shows that the lowest RMSE corresponds to 45.44 t/ha and it is achieved by using the first 23 features of the ranking generated by MDI.

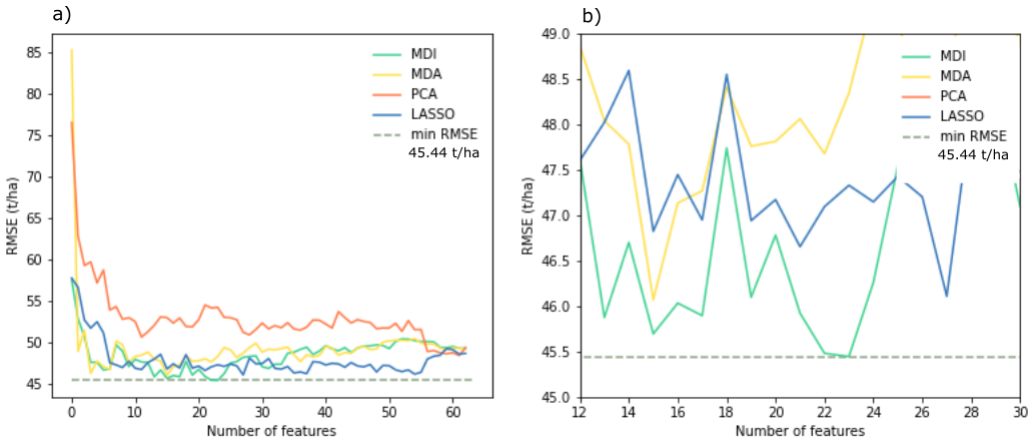


Figure 4.12: Feature selection methods testing with XGB (a) and magnification over the best performing number of features (b).

Finally, measured AGB values were plotted together with the best model predictions (Figure 4.13). For this model, values lower than 40 t/ha and higher than 160 t/ha were over- and underestimated, respectively.

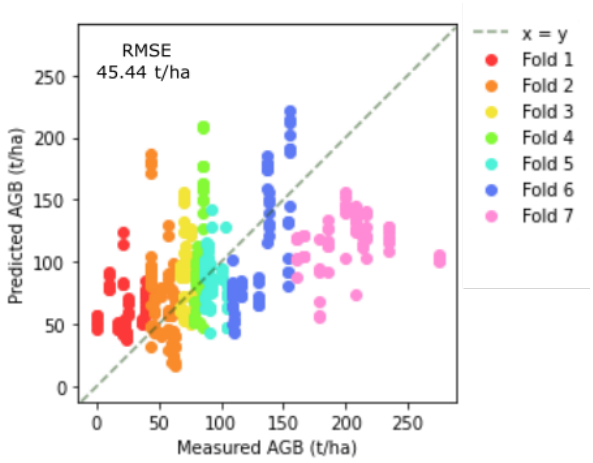


Figure 4.13: Goodness of fit on unseen data with XGB.

4.2.4 Feature selection testing with Artificial Neural Networks

The following section includes a brief description of the 3 implemented Artificial Neural Networks (ANN). Hence, a Linear (LNN), a Shallow (SNN) and a Deep Neural Network (DNN) were generated in order to test the 4 feature selection methods.

An ANN can be defined as a structure made up of densely interconnected simple processing units - also known as neurons - capable of performing considerably parallel computations (Basheer and Hajmeer, 2000). These neurons are the core processing units of the network. Hence, neurons of one layer are connected to the ones of the next layer via channels which are in charge of assigning weights. A brief and general description of the main components of these models is provided:

1. Input layer: this is constituted by the chosen features used to train the model. The inputs are multiplied by their corresponding weights and their sum is then sent to the first hidden layer, if there are any.

2. Hidden layer(s): each of the neurons belonging to this layer is associated with a value, the bias, which is added to the inputs sum received from the input layer or from a previous hidden layer. Hence, this new value is passed through an activation function which decides whether a neuron gets activated of not; therefore, activated neurons transmit the data to the next layer.

3. Output layer: the predicted output is compared to the actual output, hence the magnitude of the error is calculated, transferred backwards to our network and used for adjusting the inputs weights.

This process keeps adjusting the weights until a given number of epochs is reached; an epoch can be defined as the dataset passing forward and backward through the ANN only one time (Sharma, 2017).

Linear Neural Network

The hyper-parameters used to build the LNN are listed in Table 4.1. In order to test the 4 feature selection rankings, several LNN were trained so that each features combination could be tested.

Their performance is illustrated in Figure 4.14, a). Overall, it can been observed that the RMSE erratically decreases as new features are added into the model. The minimum error ($76.42\ t/ha$) was achieved by using the the first 39 features identified by L1 Regularization (LASSO). Additionally, the features selected by Principal Component Analysis (PCA) held the highest RMSE independently on how many features were included into the model.

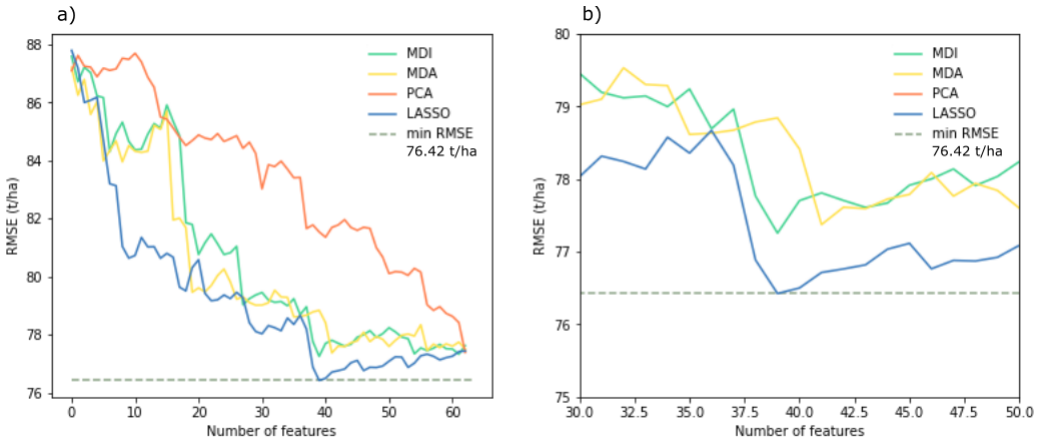


Figure 4.14: Feature selection methods testing with the Linear Neural Network (a) and magnification over the best performing number of features (b).

Linear Neural Network				
Selected variables (39 from LASSO)	LAINov, LAIcwNov, LAIcwAug, LAIcbAug, LAIcbNov, EntVIAug, DEM, ConVINov, Entb2Aug, ClreNov, b12Aug, b8Aug, NDVIre3Aug, b2Aug, b3Aug, NDVINov, EntVINov, Conb2Aug, GNDVINov, b8Nov, GNDVIAug, ConVIAug, b4Aug, NDVIre3Nov, NDVIAug, b3Nov, NDVIre2Aug, LAIAug, slope, b4Nov, MeaVINov, Meab2Nov, Meab2Aug, b6Nov, MeaVIAug, NDVIre1Aug, FAPARAugust, b2Nov, Entb2Nov			
Hidden layers	Neurons	Optimizer	Epochs	Activation Function
0	1	Adam	500	Linear
RMSE (t/ha)	76.42			

Table 4.1: Construction and performance of the Linear Neural Network

The measured AGB values were plotted together with the predictions of the best performing LNN (Figure 4.15). Hence, it can be noticed that most of the values were underpredicted.

Finally, the model was tested for over-fitting by plotting its training and validation loss. No considerable over-fitting was found for the LNN, as visible in Figure 4.16.

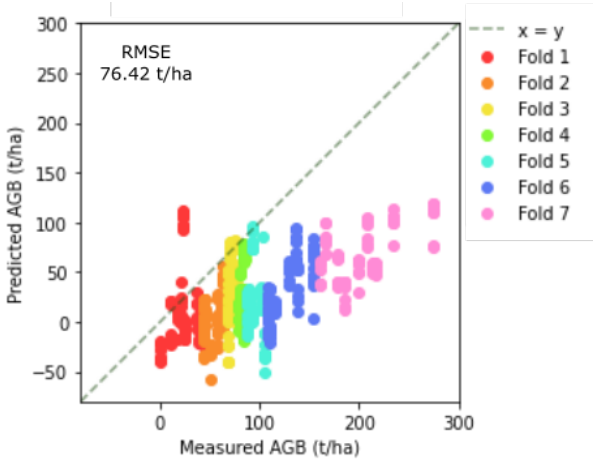


Figure 4.15: Goodness of fit on unseen data with the Linear Neural Network.

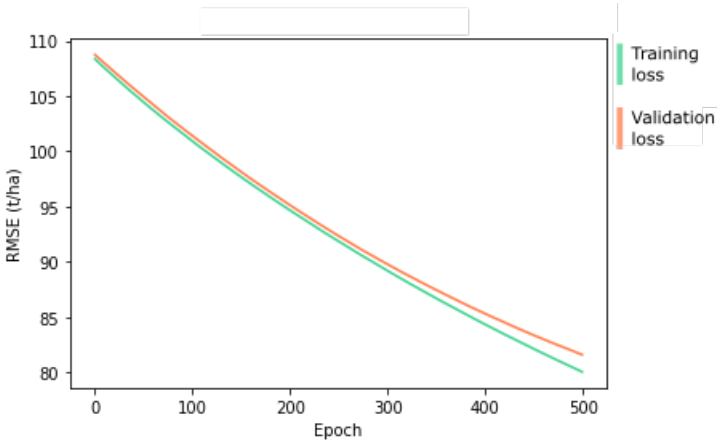


Figure 4.16: Training and Validation loss in the Linear Neural Network.

Shallow Neural Network

A Shallow Neural Network (SNN) was built by using the hyper-parameters listed in Table 4.2. This is composed of 2 hidden layers made up of 8 and 4 neurons; hence, a rectified linear activation function (Relu) was used for the inner layers. This function works by directly outputting the input value when this is positive and outputting zero when the it is negative. Finally, a linear function is used for the last layer.

In order to evaluate the feature selection rankings, several SNN were trained so that each features combination could be tested. Their performance is illustrated in Figure 4.17, a). Overall, the RMSE showed a very erratic pace, and reached a minimum (42.62 t/ha) when 15 features identified by LASSO were added to the model.

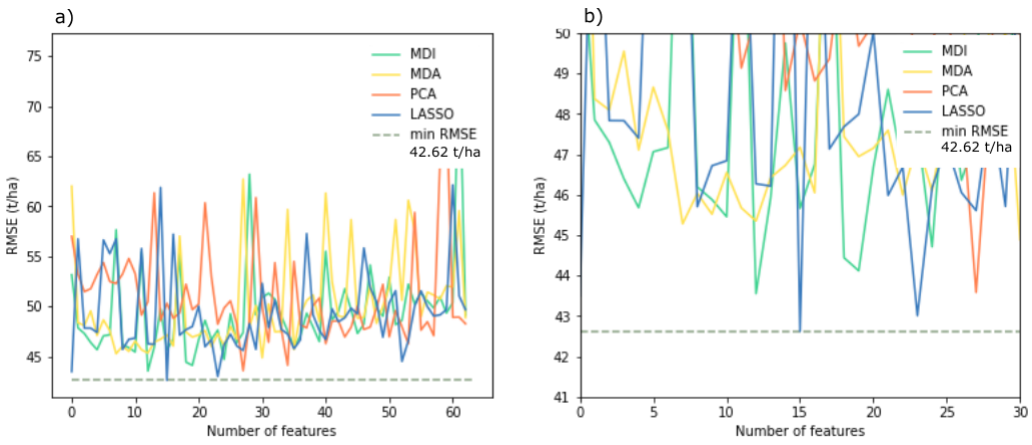


Figure 4.17: Feature selection methods testing with the Shallow Neural Network (a) and magnification over the best number of features (b).

Shallow Neural Network				
Selected variables (15 from LASSO)	LAINov, LAIcwNov, LAIcwAug, LAIcbAug, LAIcbNov, EntVIAug, DEM, ConVINov, Entb2Aug, ClreNov, b12Aug, b8Aug, NDVIre3Aug, b2Aug, b3Aug			
Hidden layers	Neurons	Optimizer	Epochs	Activation Function
2	8, 4, 1	Adam	500	Relu & Linear
RMSE (t/ha)	42.62			

Table 4.2: Construction and performance of the Shallow Neural Network

Additionally, the measured AGB values were plotted together with the prediction of the SNN yielding to the lowest RMSE (Figure 4.18). Hence, it can be noticed that values lower than 40 t/ha and higher than 160 t/ha were over- and under-predicted, respectively.

Finally, the SNN was tested for over-fitting (Figure 4.19). It can be observed that, as the number of epochs increased, the training and validation loss started diverging; this indicates that the model was over-fitting the training data.

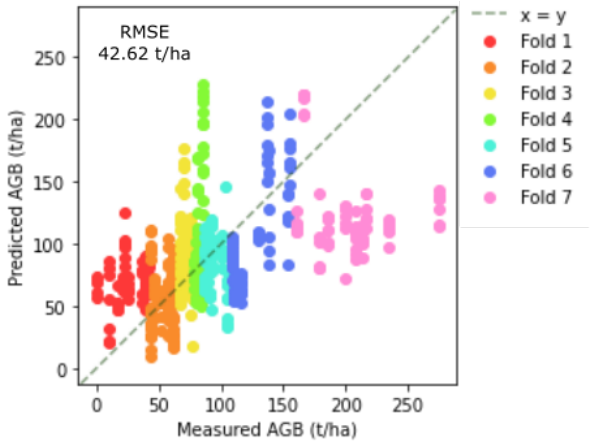


Figure 4.18: Goodness of fit on unseen data with the Shallow Neural Network.

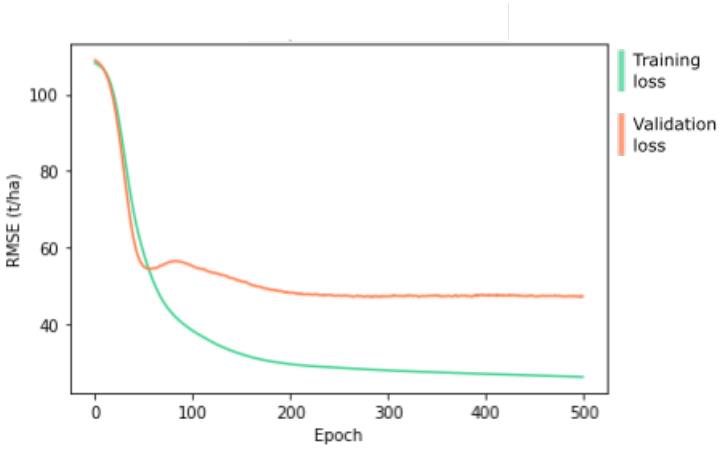


Figure 4.19: Training and Validation loss in the Shallow Neural Network.

Deep Neural Network

The hyper-parameters set to build the Deep Neural Network (DNN) are listed in Table 4.3. This has 5 hidden layers made up of 128, 64, 32, 16 and 8 neurons; hence, a Relu activation function is used for the inner layers, while a linear function is set for the output layer. Moreover, 2 dropouts are added. Briefly, by setting a dropout the network ignores certain neurons during the training phase; this is an approach to Regularization in ANNs, and is used to prevent over-fitting in Deep Neural Networks (DNN) (Srivastava et al., 2014).

With the scope to evaluate the feature selection rankings, several DNN were trained so that each features combination would be tested. Their performance is illustrated in Figure 4.20, a). Overall, it can been observed that the RMSE had a very erratic pace, and after and initial decrease, it started increasing after more than circa 5 features were added. The best performing DNN was achieved when using the 5 features identified by MDI, which yielded an error of 42.30 t/ha.

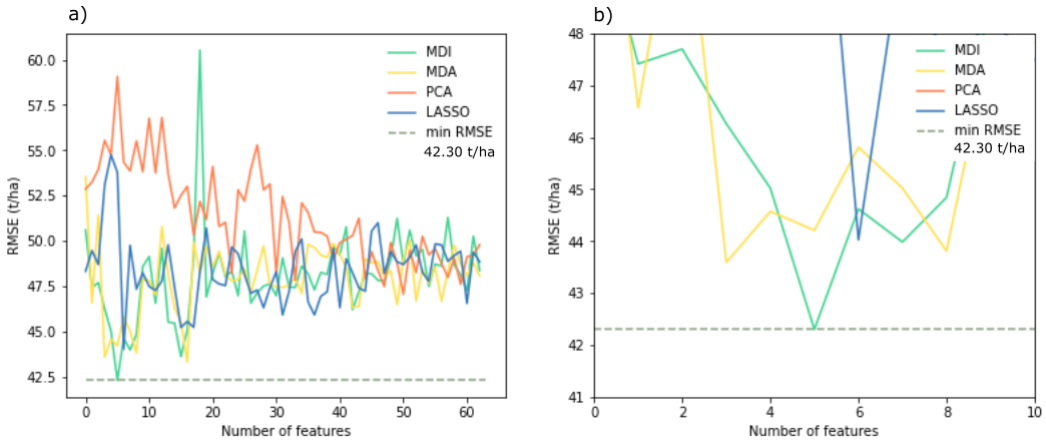


Figure 4.20: Feature selection methods testing with the Deep Neural Network (a) and magnification over the best performing number of features (b).

Deep Neural Network				
Selected variables (5 from MDI)	LAIcbAug, LAIcwAug, ClreAug, b12Aug, DEM,			
Hidden layers	Neurons	Optimizer	Epochs	Activation Function
5 + 2 dropouts*	128, 64, 32, 16, 8, 1	Adam	500	Relu & Linear
RMSE (t/ha)	42.30			

**A dropout of 0.3 is found between the first and second hidden layer, and another dropout of 0.1 is found between the third and the fourth hidden layer.*

Table 4.3: Construction and performance of the Deep Neural Network

Therefore, the unseen AGB values were plotted together with the prediction of the DNN yielding to the lowest RMSE (Figure 4.21). Hence, it can be noticed that values lower than 40 t/ha and higher than 160 t/ha were over- and under-predicted, respectively.

Finally, the DNN was tested for over-fitting (Figure 4.22). It can be observed that the training and validation loss diverged and the first was over-fitting the training data. Hence, the validation loss stayed constant throughout the epochs.

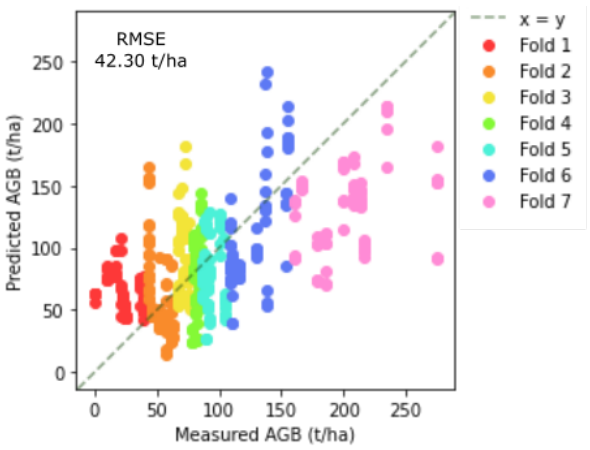


Figure 4.21: Goodness of fit on unseen data with the Deep Neural Network.

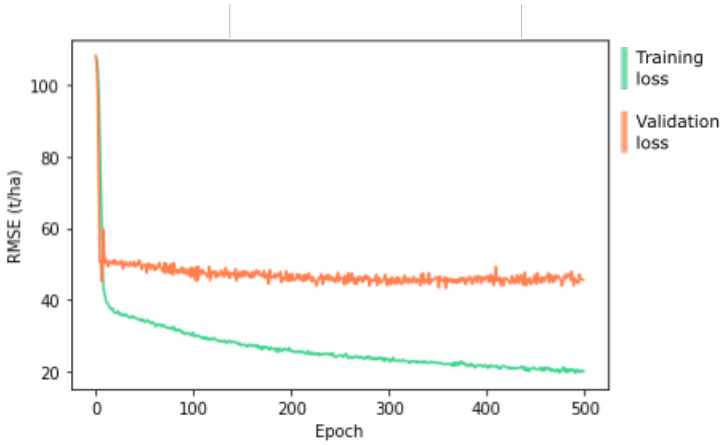


Figure 4.22: Training and Validation loss in the Deep Neural Network.

Overall, all of the tested models selected different feature selection methods, without showing any evident preference; an exception was found for the PCA method, which yielded to the worst predictions for every model. Specifically, the feature ranking generated by the MDI method resulted in lower RMSE for XGB and DNN. These features mostly included biophysical parameters from the Summer month, a Short-wave infrared band (band 12), a chlorophyll index derived from the red-edge bands and topographic variables. The XGB, which included more features, was modelled using also biophysical parameters from the Autumn month and texture measures. The MDA method, which was chosen by RF and kNN, ranked the features is a similar way to MDA; the best performing included topographic variables, biophysical parameters from the Summer month, band 12, and a chlorophyll index derived from the red-edge bands. Finally, LASSO was chosen by the LNN and the SNN; this method identified, as most important features, biophysical parameters and texture measures from August and November, band 2, 3, 8 and 12 and Vegetation Indices derived from the red-edge bands.

4.3 Hyper-parameters fine-tuning

L1 Regularization (LASSO) for feature selection was introduced in Section 4.1.2. Such a model attempts to find the values for the weights $w_0, w_1, \ldots, w_n$ which minimize the error given by the formula 4.4. The way in which weights are optimized is heavily dependent on the value of alpha (α), which is not being tuned by the model itself, but it is rather predefined by the user. Therefore, α is what is known as a hyper-parameter.

$$error = \|y - X_0 w_0 + X_1 w_1 + X_2 w_2 + \ldots + X_n w_n\| + \underbrace{\alpha \sum |w_i|}_{regularization\ term} \tag{4.4}$$

Other examples include the number of trees in a Random Forest (RF), the number of neighbours in K-Nearest Neighbour (kNN), as well as the learning rate of an Artificial Neural Network (ANN). Therefore, in order to improve the model performance, the user must optimize its hyper-parameters by following different approaches. The following section illustrates 3 hyper-parameters optimization methods - Grid Search, Random Search and Bayesian Search - for 3 of the previously tested algorithms - KNN, RF, and Extreme Gradient Boosting (XGB). Hence, a range of values for each tested hyper-parameter was predefined by the author before any search method was applied.

Hyper-parameters optimization with Grid Search involves the generation of many models, so as to test all the possible hyper-parameters combinations. This approach always yields to the combination resulting in the best performing model; on the other hand, when the problem is multidimensional - i.e. more than one hyper-parameter needs to be fine-tuned - the computation can become extremely time-consuming. A visual explanation of hyper-parameter fine-tuning using the Grid Search method is found in Figure 4.23, where a 1-dimensional problem is illustrated.

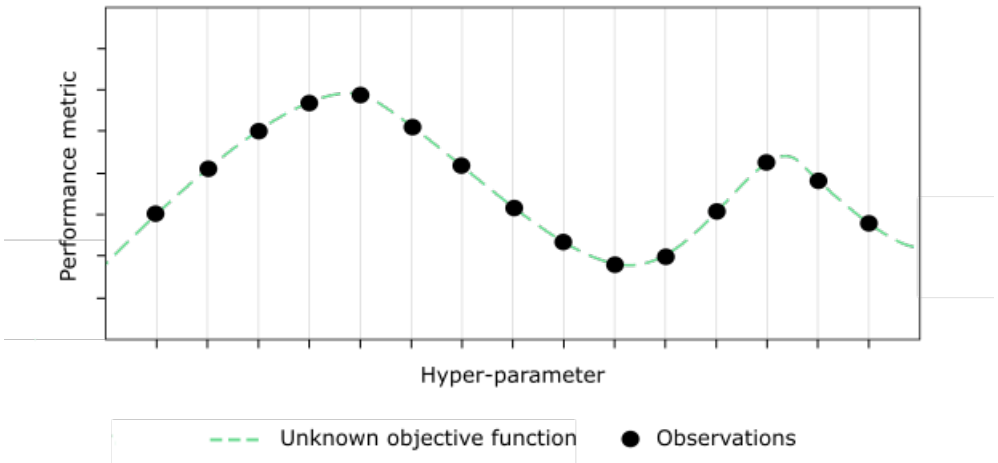


Figure 4.23: Grid Search with a 1-dimensional problem.

Random Search looks for the best hyper-parameters combination by creating as many models as the number of predefined iterations. The success or failure of this approach in increasing the model performance is totally random and it cannot give the user any guarantees; on the other hand, the computation is very fast if the number of iteration is kept low. A visual explanation of hyper-parameter fine-tuning using the Random Search method is found in Figure 4.24, where a 1-dimensional problem is illustrated.

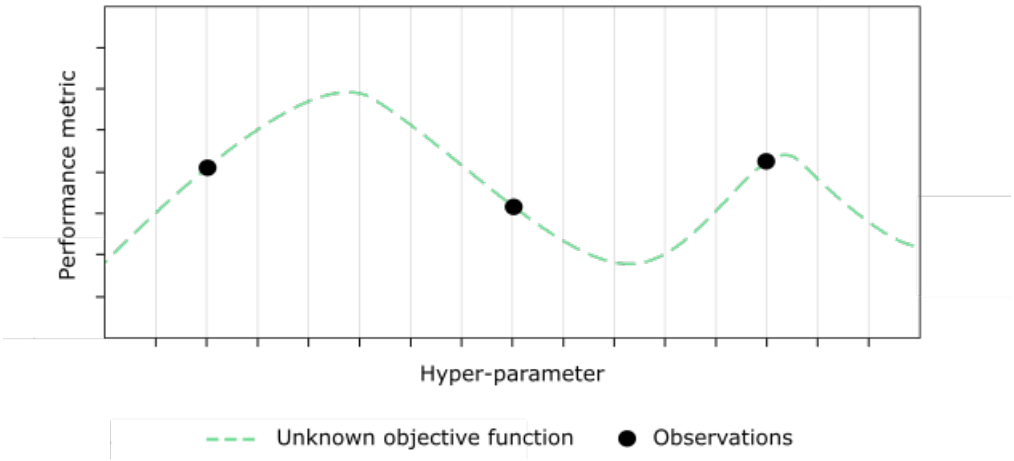


Figure 4.24: Random Search with a 1-dimensional problem.

Bayesian optimization was implemented by using a python package created by Nogueira (2014); who defines this algorithm as a constrained global optimization tool which needs a performance metric to be maximized in as few iterations as possible, through Bayesian inference and Gaussian process. For this work, the negative RMSE (-RMSE) was set as the value to be maximised. Subsequently, the hyper-parameters which want to be tested for each algorithm and their respective ranges were identified. Hence, Bayesian optimization builds a probabilistic model for the selected performance metric and search for the hyper-parameters combination which maximises the value of such a chosen metric.

Hyper-parameters optimization with Bayesian Search has been proved useful for high-dimensional problems (Snoek et al., 2012); this approach is here used as an alternative to the long computational time characterizing a Grid Search and to the pure randomness which we deal with when using Random Search.

A visual explanation of hyper-parameter fine-tuning using the Bayesian optimization method is found in Figure 4.25, where a 1-dimensional problem is illustrated. In this example, we have a certain performance metric to be maximised, and one only hyper-parameter that is going to be tested and for a given range. Hence, the algorithm is initialized with 5 randomly chosen values for the hyper-parameter. Next, the optimization process starts and it can be divided into 3 major steps: updating the probability model of the loss function; choosing a new hyper-parameter value (next best guess); and evaluating the model performance for that new value.

The probability model of the loss function - or prediction objective function - is evaluated by using Bayesian inference - a method used for updating a probability once new information is available - and Gaussian process. The Gaussian Process is the generalization of

a Gaussian distribution where, instead of over random numbers, the distribution is over a function; furthermore, just like the Gaussian distribution is defined by its mean and its variance, a Gaussian Process is defined by its mean function and its standard deviation function (Brochu et al., 2010). Hence, for every point in our prediction function, the Gaussian Process returns the mean and standard deviation of the normal distribution at that point. Figure 4.25 shows some areas with known values around which the uncertainty of the prediction is very low, whereas areas in which no value was sampled the uncertainty of the prediction is very high.

Once the initial probability model is generated, a new value for the hyper-parameter is tested. Bayesian optimization does this thanks to an utility function (or acquisition function) (Snoek et al., 2012). High values for the acquisition function are found in regions where the Gaussian Process predicted high performance (exploitation zone) or in regions where uncertainty is very high (exploration zone); areas characterized by both exploration and exploration are sampled first as next best guess (Brochu et al., 2010). Once the new hyper-parameter value is sampled, the model performance is evaluated, and the probability model is updated with the new information; this cycle is repeated until a certain threshold is reached or once the iterations are over.

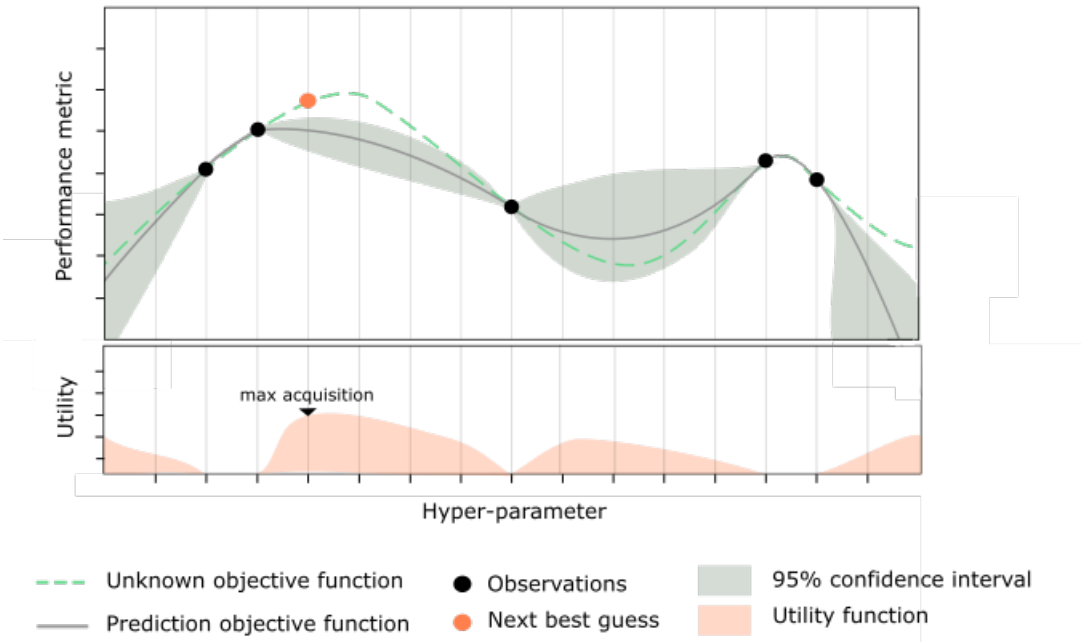


Figure 4.25: Bayesian Search with a 1-dimensional problem.

K-Nearest Neighbour fine-tuning

The Grid Search hyper-parameter optimization for the kNN algorithms was performed by using as model input the 25 previously selected variables from the MDA feature selection method, and by fine-tuning the number of neighbours (k) is order to find the best performing kNN model. As showed in Figure 4.26, a range of values from 0 to 50 were tested. By observing the left-hand-side of the graph, it can be noticed that the RMSE

plummet when higher number of neighbours was added, until reaching 11 neighbours, which produced the lowest validation error. Furthermore, as more neighbours are added the performance of the model starts decreasing again.

The best performing kNN model was finally defined; this yielded an error of 42.68 t/ha, which was achieved by using as independent variables the 25 features selected by the MDA measure, and 11 neighbours (Table 4.4).

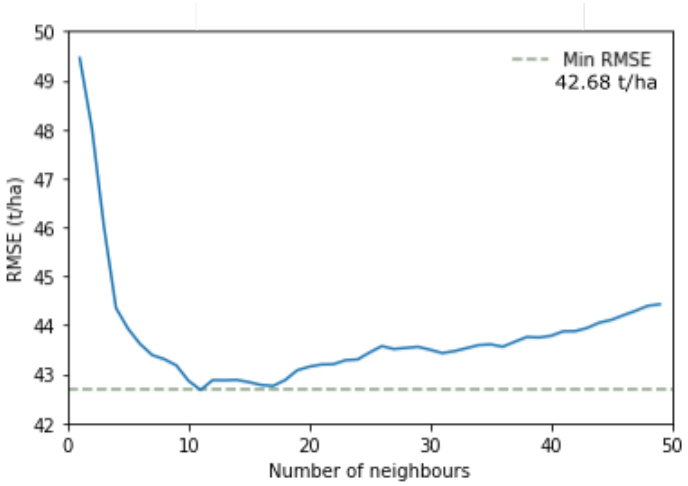


Figure 4.26: KNN optimization with Grid Search.

K-Nearest Neighbour		
Selected variables (25 from MDA)	DEM, LAIcbAug, b12Aug, LAIcwAug, ClreAug, EntVIAug, FAPARAug, FAPARNov, slope, GNDVIAug, LAINov, LAIcbNov, Conb2Aug, FCOVERNov, MeaVINov, LAIAug, Entb2Aug, EntVINov, LAIcwNov, Entb2Nov, b3Aug, b11Aug, GEMINov, MeaVIAug, Meab2Aug	
Hyper-parameter	Default	Grid Search
Number of Neighbours	5	11
RMSE (t/ha)	43.93	42.68

Table 4.4: K-Nearest Neighbour model construction.

Random Forest fine-tuning

Hyper-parameters optimization for Random Forest (RF) was performed by using as model input the 3 previously selected variables from the MDI feature selection method, and by fine-tuning 4 hyper-parameters within their respective range of values: maximum depth (from 10 to 100), minimum sample leaf (from 1 to 2), number of estimators (from 30 to 1000), and the minimum sample split (from 2 to 5). Hence, the performance reached by each optimization approach is showed in Table 4.5.

It can be observed that the Grid Search method yielded the lowest validation error (44.13 t/ha) amongst the tested RFs, however it also required over 96 hours of computational time. On the other hand, the Random Search was the worst hyper-parameters fine-tuning method with a RMSE of 44.26 t/ha. Overall, all the optimization approaches yielded a better performance compared to using default hyper-parameters; however, it should be noticed that such differences in errors are still too small to be considered relevant.

Random Forest				
Selected variables (3 from MDA)	DEM, LAIcbAug, b12Aug			
Hyper-parameters	Default	Grid Search	Random Search	Bayesian Search
Max depth	none	19	30	24
Min samples leaf	1	1	1	1
Number of estimators	100	95	140	96
Min samples split	2	2	2	2
RMSE (t/ha)	44.29	44.13	44.26	44.16

Table 4.5: Random Forest model construction.

Extreme Gradient Boosting fine-tuning

Hyper-parameters optimization for XGB was performed by using as model input the 23 previously selected variables from the MDI feature selection method, and by fine-tuning 6 hyper-parameters within their respective range of values: Alpha (from 0 to 0.9), Gamma (from 0 to 0.9), learning rate (from 0 to 0.9), Number of estimators (from 30 to 1000), maximum depth (from 1 to 8), and the sub-sample (from 0 to 0.9). Hence, the performance reached by each of the tested optimization approaches is illustrated in Table 4.6.

Due to its long computation time, Grid Search was not tested for the XGB algorithm. However, an important decrease in error was reached by implementing Bayesian optimization (37.79 t/ha). On the other hand, the Random Search was again the worst hyper-parameters fine-tuning method with a RMSE of 94.34 t/ha. Overall, not all the optimization approaches yielded a better performance compared to using default hyper-parameters.

Extreme Gradient Boosting			
Selected variables (23 from MDI)	LAIcbAug, LAIcwAug, ClreAug, b12Aug, DEM, FAPARAug, EntVIAug, GNDVIAug, FAPARNov, slope, LAINov, LAIcbNov, LAIAug, Conb2Aug, FCOVERNov, MeaVINov, EntVINov, Entb2Aug, LAIcwNov, GEMINov, ClreNov, Entb2Nov, b4Aug		
Hyper-parameters	Default	Random Search	Bayesian Search
Alpha	0	0	0.05
Gamma	0	0	0.30
Learning rate	0.30	0	0.40
Number of estimators	100	178	130
Max depth	6	1	1
Subsample	1	0	0.7
RMSE (t/ha)	45.44	94.35	37.79

Table 4.6: Extreme Gradient Boosting model construction.

4.4 Above Ground Biomass and Carbon Sequestration mapping

The following section contains a usual comparison amongst Above Ground Biomass (AGB) maps generated by using the prediction of 3 different models. Hence, the final AGB map and related Carbon Sequestration (CS) map are presented.

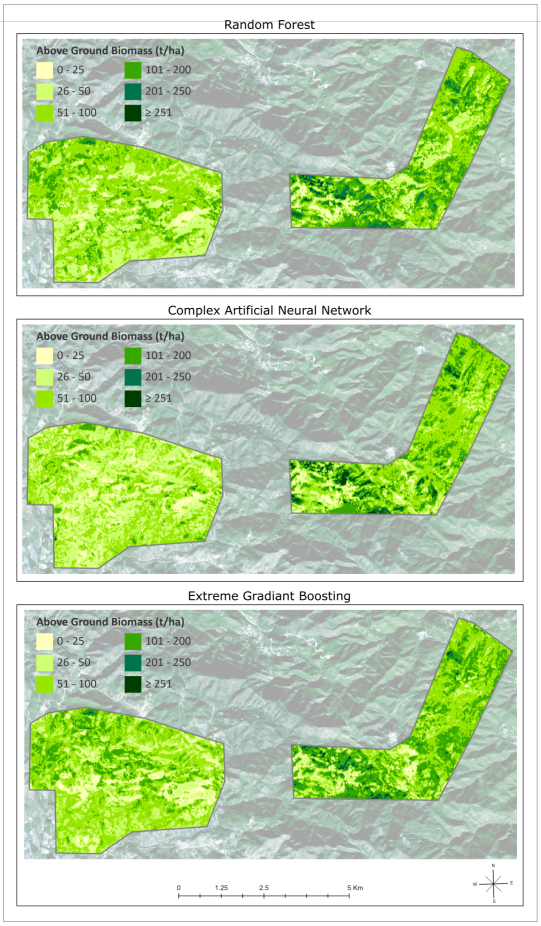


Figure 4.27: Comparison between AGB maps generated from different models.

Amongst the tested RF models, the one which yielded the best performance was used to produce the first AGB map of Figure 4.27. This model was built by using the 3 top-ranked features selected by the MDA measure, and hyper-parameters optimization with the Grid Search. Such a model led to a validation error of 44.16 t/ha. The estimated AGB ranges from 0.72 t/ha to 274 t/ha, with an average value of 81.31 t/ha and a standard deviation of 52.14 t/ha (Figure 4.28). The second map of Figure 4.27 was generated by using the Deep Neural Network (DNN) predictions, which yielded to the best performance amongst the tested ANN. This model was built by using the 5 top-ranked features selected by the MDI measure. Such predictions led to a validation error of 42.30 t/ha. The estimated AGB ranges from 9.18 t/ha to 500 t/ha, with an average AGB value of 76.69 t/ha, and a standard deviation of 56.87 t/ha (Figure 4.28). The final map depicted in Figure 4.27

was generated by using the prediction from the XGB model which yielded the lowest error compared to all of the previously tested algorithms. This model was built by using the 23 top-ranked features selected by the MDI measure, and hyper-parameters optimization with the Bayesian Search. The model led to a validation error of 37.79 t/ha, an estimated average AGB value of 83 t/ha, a minimum of 0 t/ha, a maximum of 346.56 t/ha and a standard deviation of 51.3 t/ha (Figure 4.28).

Overall, the predictions from all of the mentioned models as well as the field-based AGB values resulted in distributions skewed to the right, with a higher frequency of low to average values. The DNN predictions resulted in a lower average value compared to the other models.

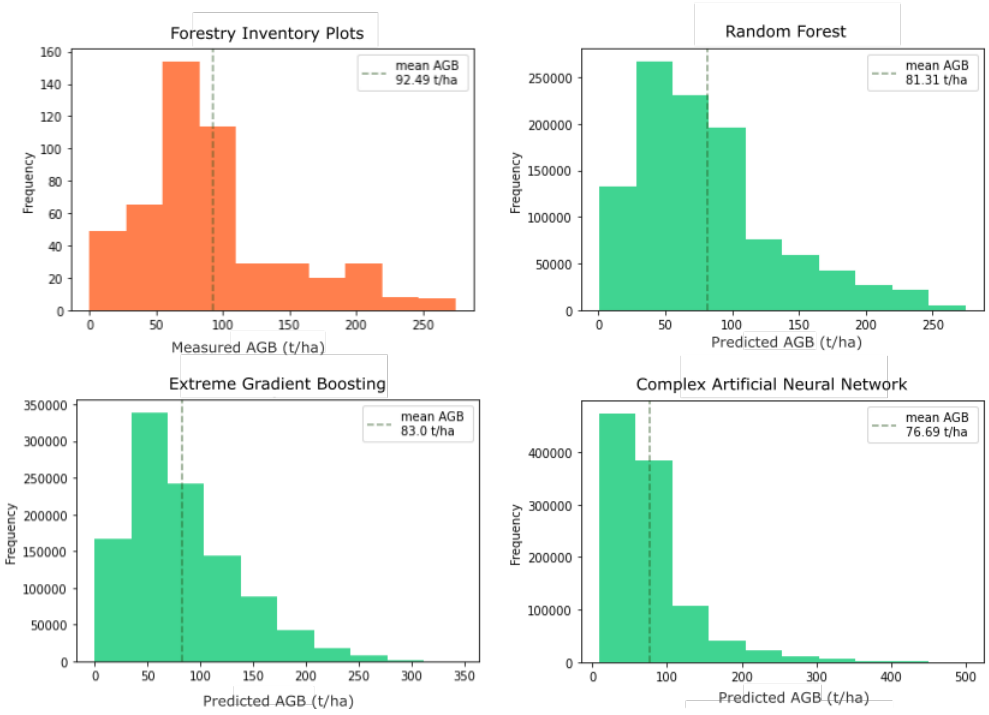


Figure 4.28: Comparison between AGB predictions generated from different models.

The XGB model was used for the generation of the final AGB and CS maps. The AGB map in Figure 4.29 was generated by using 6 manual classes; it can be observed that the majority of the territory located on the south-west of map is covered by AGB values ranging from 0 t/ha to 100 t/ha, with spread high values ranging from 101 t/ha to 200 t/ha. On the other hand, the second portion of the study area, located on the north-east side of the map, is characterized by higher AGB values, with areas reaching over 250 t/ha.

Finally, a Carbon Sequestration (CS) map was generated by using the conversion factor suggested by the IPCC (2006), that is multiplying the predicted AGB values for 0.47. The CS map in Figure 4.30 was generated by manually defining 6 classes; the portion of the study area located on the south-west of map is sequestrating 0 t/ha to 40 t/ha, with spread higher values ranging from 41 t/ha to 80 t/ha. On the other hand, the second portion of the study area, located on the north-east side of the map, is characterized by higher CS, with areas storing over 120 t/ha.

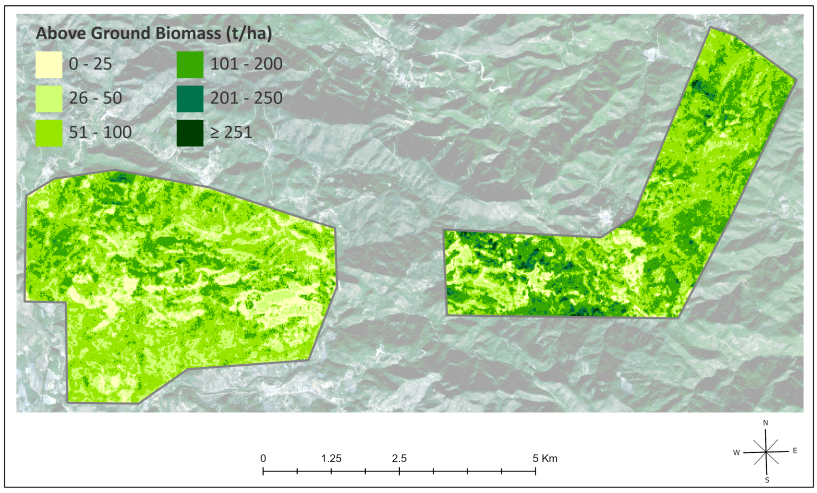


Figure 4.29: Map of estimated Above Ground Biomass.

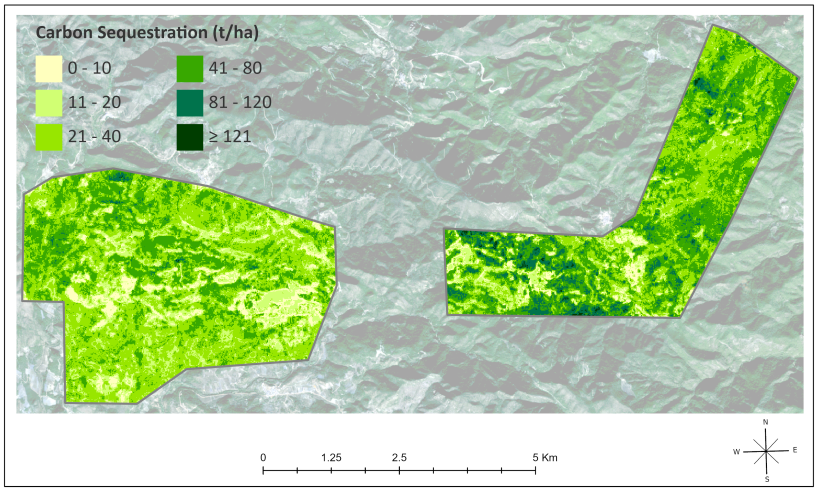


Figure 4.30: Map of estimated Carbon Sequestration.

4.5 Model explanation with Shapley values

Ideally, a Machine Learning (ML) model should be highly accurate and simple to interpret. By "simple to interpret" is intended the ability to expose its performance in an understandable and intuitive way. Unfortunately, the more the model complexity increases, the harder it is to understand how certain values were predicted and which features had contributed more to those predictions. When working with a simple model, the impact that a feature has on the model output is easily interpretable by looking at its weights; whereas, complex models such as ensemble methods or deep networks are not as easy to understand; in this scenario, a model explainer can help interpreting the model results (Lundberg and Lee, 2017).

In the following section, the SHapley Additive exPlanations (SHAP) package was used as model explainer for the best performing Extreme Gradient Boosting (XGB) model. SHAP was created by Lundberg et al. (2020) [1] and, according to its author, its implementation allows for appropriate user trust, provides insights for model improvement, and supports the understanding of the problem being modelled (Lundberg and Lee, 2017). It should be noticed that only the 20 first most influential features can be visualised by the SHAP package.

This approach differs from the more common feature importance methods such as the previously explained MDA, which solely compute the global feature importance. Whereas, SHAP is a local method, meaning that a local feature importance is calculated for every observation. This is performed by holding out the feature value before predicting each instance; which is then repeated for each feature and each instance of the entire training set so as to measure the local importance of each feature. The computation of each instance for big dataset can become time consuming, however, in 2020 a fast and precise algorithm was created for tree-based models (Lundberg et al., 2020).

A global feature importance can be obtained by aggregating the local feature importance of each instance, as shown in the summary plot in Figure 4.31, where each dot represents an observation of the dataset. This plot illustrates how the model predictions were influenced by each feature. More specifically, it shows the features contribution, for each instance, in pushing the model output from a base value to the output value; where the base value is defined as the average model output over the training set (Lundberg et al., 2020). When a feature has negative SHAP value, the dot representing that instance is found on the left-side of the plot and it means that that feature value pushed the prediction for that instance (or dot) to be lower than the base value; on the other hand, if a dot is found on the right-hand side of the plot, the observed feature value pushed the model output to be higher than the base value. Furthermore, for each feature, overlapping points are visualized in the y-axis direction so to give an idea of the feature values distribution. Additionally, the colour of the dots refers to the features value.

Finally, by observing the summary plot in Figure 4.31 the features can be divided into 2 groups: 1) features which increasing in values pushed the model to output AGB higher than the base value; 2) features which increasing in values would push the model to output AGB lower than the base value. Belonging to the first group are: The Chloro-

[1] `https://github.com/slundberg/shap`

phyll index based on red-edge bands from the Summer month (ClreAug), the Fraction of absorbed radiation from both months (FAPARAug and FAPARNov), the Canopy Chlorophyll Content from the Summer month (LAIcbAug), and the slope; while, band 12 from the Summer image (b12Aug) and the DEM belong to the second group. The remaining features do not show any clear visual pattern in the way they impacted the model output.

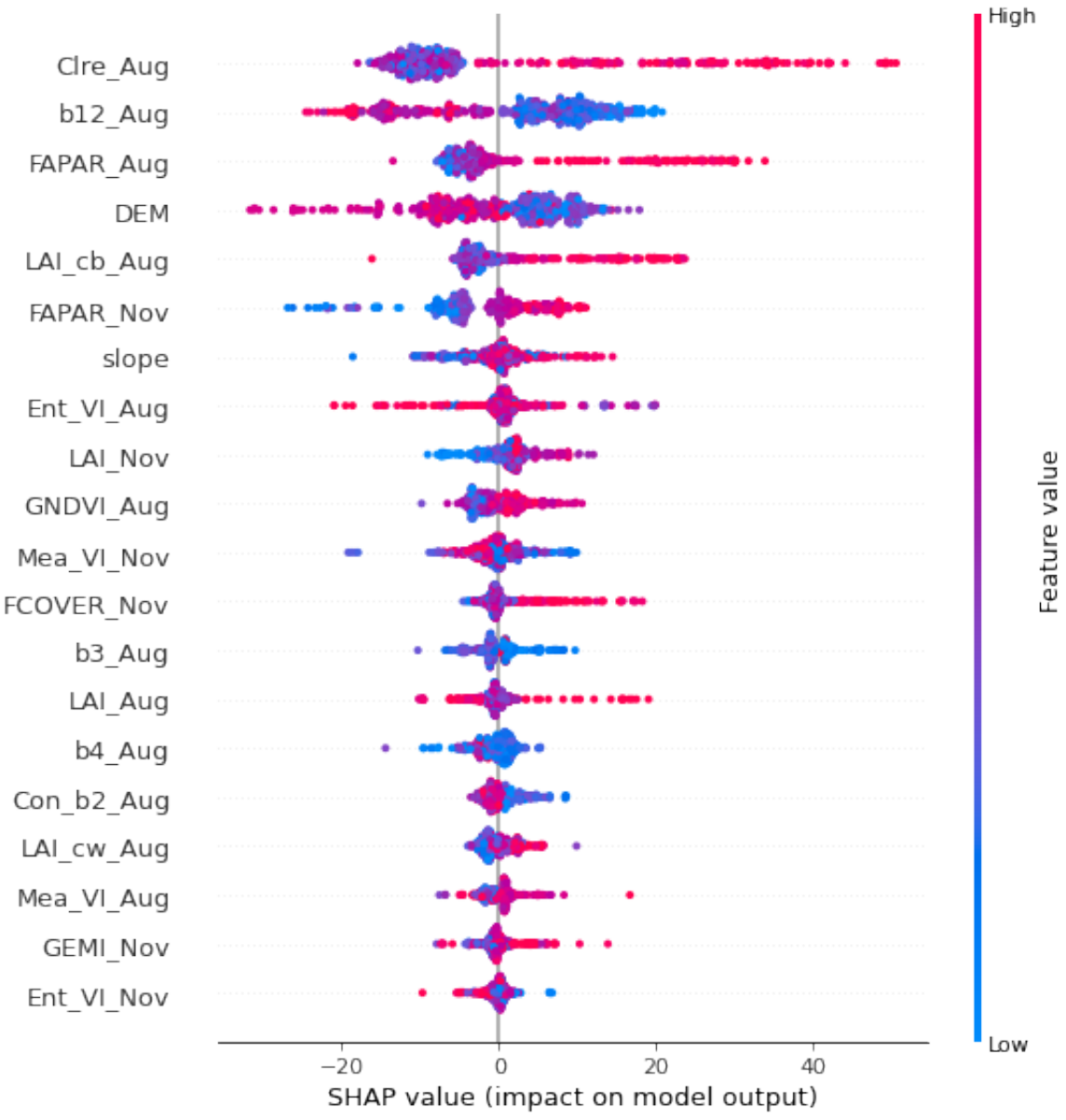


Figure 4.31: Summary plot with SHAP.

Chapter 5

Discussion

The presented work aimed to assess the potential of Sentinel-2 for Above Ground Biomass (AGB) estimation by exploring a broad palette of methods from the domain of Machine Learning (ML). The following Chapter discusses the findings, strengths and limitations of this work and it is structured in the form of answers to the Research Questions (RQ) stated in Chapter 1.

The 1st and 2nd RQs sought to identify an optimal feature selection method and which of these features would yield to an optimal model performance, respectively. This work tested 4 feature selection methods: Mean Decrease in Impurity (MDI), Mean Decrease in Accuracy (MDA), Principal Component Analysis (PCA) and L1 Regularization (LASSO). Hence, their performance was tested by several regression models: K-Nearest Neighbour (K-NN), Random Forest (RF), Extreme Gradient Boosting (XGB), a Linear Neural Network (LNN), a Shallow Neural Network (SNN) and a Deep Neural Network(DNN).

The features ranking resulted from MDI, MDA and LASSO were chosen equally often by the models, resulting in none of these methods clearly standing out. Whereas, PCA gave the worst performance for every model. Being PCA an unsupervised algorithm based on variance, we can say that none of the models benefited from using features which, despite explaining a high percentage of variance in the dataset, were not necessarily correlated with the target. The ranking provided by MDA and MDI resulted to be very similar; this is sustained by the fact that both of them are measures derived from the RF algorithm. Moreover, RF and XGB chose the features ranking by MDA and MDI, respectively. Therefore, tree-based models performed better with RF derived features rankings. The two methods (MDA and MDI) are advantageous as they are of easy implementation, with MDI having a slightly faster computational time. Moreover, despite LASSO ranked the features in a different order compared to MDI and MDA, it showed a very similar performance in terms of model accuracy and quick computation time. Also, despite its implementation not being very straightforward, feature selection with LASSO should be further explored especially when working with models which are not tree-based - i.e. Artificial Neural Networks (ANN).

Overall, Summer features were selected more often than Winter features by all feature selection methods. Out of the 63 initially generated variables, only 23 features selected by MDI were used for the final model. The highest ranking features by their importance are: biophysical parameters (LAI, LAI cw, FAPAR) from the Summer month, the Digital

Elevation Model (DEM), a SWIR band(band 12) and a Chlorophyll index generated from the red-edge bands (Clre), both from the Summer month.

Still, there was a significant difference among the number of selected features, with the lowest number (n = 3) selected for the RF by MDA, and the highest number (n = 39) selected for the LNN. This fact, together with the fact that there was not a clear winner among the feature selection methods, suggest that regression models may be expressing preference to a specific feature selection method. Indeed, the models tended to need a lesser number of features when these were ranked by MDI and MDA.

In the context of AGB estimation, several authors have been performing feature selection based on either stepwise regression or RF derived measures (Gao et al., 2018; Lu, 2006). However, further studies including more feature selection methods testing (such as Monte Carlo simulations) should be performed in order to confirm or deny the hypothesis of preference towards any methods.

The 3rd and 4th RQs aimed to identify the optimal hyper-parameters fine-tuning method and regression model, respectively, for the estimation of AGB in the study area. The following hyper-parameters optimization methods were tested: Random, Grid and Bayesian Search. Hence, the Grid Search provided the lowest errors for the tested models. However, due to its long computation time, it was inconvenient for it to be tested in models requiring many hyper-parameters. Moreover, the decrease in error achieved when using Grid and Random Search, was small and irrelevant, as this was only a few decimal points. On the other hand, Bayesian Search significantly improved the XGB model predictions; this Search method had advantageous computation time compared to the Grid Search, yet its implementation was slightly more complex.

The models with high number of free hyper-parameters exhibited the highest improvement of predictive performance after hyper-parameter fine-tuning. Therefore, we can say that the importance of hyper-parameters optimization depended on the model used. In this work, the XGB was found to be the best performing algorithm when used in combination with MDI and Bayesian Search. The importance of hyper-parameter fine tuning for models with many hyper-parameters was confirmed by Pandit et al. (2018); however, the manual hyper-parameters optimization implemented in their study improved the performance of only a few irrelevant decimal points.

For all the tested models, AGB values lower than 40-50 t/ha were slightly overpredicted whereas values higher than 150-160 t/ha were underpredicted. As it was confirmed by previous studies (Galidaki et al., 2017; Salem et al., 2020; Forkuor et al., 2020), this is a typical issue when estimating AGB with the use of ML and satellite images, and it is exacerbated by a limited number of representative samples for low and high values of the forestry inventory.

Answering to the 5th RQ allowed for collecting more insights on the input features and their influence on the model output. This was achieved thanks to a model explainer, that was the SHapley Additive exPlanations (SHAP) package. It was found that high values of a Vegetation Index generated from the red-edge bands and biophysical parameters extracted from the Summer season, lead to an increase in predicted AGB. The same positive behaviour was found for certain biophysical parameters extracted

from the Autumn Season. In contrast, high values of the Summer SWIR band and high elevation pushed the predicted AGB to values lower than the baseline. Those findings are in agreement with the scientific consensus on the relationship between elevation and biomass.

The insights gathered with SHAP showed the utility of a model explainer for the scientific community. Further work should be focusing on the exploration of model explainers so as to allow for appropriate user trust, to provide insights for model improvement, and to support the understanding of the problem being modelled (Lundberg and Lee, 2017). Hence, a model explainer is a beneficial tool to use when communicating the study results to clients.

Finally, by comparing different methods for each of the main methodological cornerstones (Figure 3.4), this work contributed to our understanding of the limitations and strengths of those approaches and underlines the importance of feature selection and well as of model optimization and explainability.

Chapter 6

Conclusion

This work implemented a Sentinel-2 based exploratory workflow for the estimation of Above Ground Biomass (AGB) and Carbon Sequestration (CS). In the last decades, remote sensing-based studies on AGB have been widely investigated. Up-to-date mapping of such measures are required by international commitments under the climate convention as well as by sustainable forest management practices. Hence, this study explored a palette of feature selection methods, regression models, hyper-parameters optimization approaches, and a model explainer. We conclude that:

- Sentinel-2 images - especially when collected in the Summer month - effectively predicted AGB. The most influential features were found to be: a Vegetation Index (VI) derived from red-edge bands (Clre Aug), a SWIR band (b12 Aug), and biophysical parameters (FAPAR Aug, LAI cb Aug, FAPAR Nov). Also, the Digital Elevation Model (DEM) showed strong influence on the predictions.
- Amongst the 4 tested feature selection methods - Mean Decrease in Impurity (MDI), Mean Decrease in Accuracy (MDA), L1 Regularization (LASSO) and Principal Component Analysis (PCA) -, MDI, MDA and LASSO led to similar results in terms of models performance; therefore, the easiness of use and fast speed of MDI makes it more advantageous. It is unclear which feature selection method performed the best, this will need further studies and depends on the model used. On the other hand, the PCA method was constantly unsuccessful.
- All of the tested regression models - k-Nearest Neighbour (kNN), Random Forest (RF), Extreme Gradient Boosting (XGB), Linear Neural Network (LNN), Shallow Neural Network (SNN), and Deep Neural Network (DNN) - benefited from feature selection. After this step, the lowest error was reached by the DNN with the first 5 features ranked with the MDI.
- The hyper-parameters optimization step with Bayesian search was beneficial. While Grid search was very time-consuming, and Random search was too unpredictable, Bayesian search proved to be an interesting tool for improving the performance of the XGB model.
- Of all the tested combinations, the most reliable ensemble for predicting AGB in our study area was found to be: MDI for feature selection, XGB as regression model, and Bayesian Search for hyper-parameters optimization.

- Finally, the insights that this study gathered by using the SHAP package, showed its potential utility for the community. However, a single prospective is never enough in the scientific community; therefore, further work should be focusing on the exploration model explainers so as to support a deeper understanding of the problem being modelled, to gather insights on how that can be improved and it can also be used as a straightforward means to communicate results to clients.

Bibliographic References

Altmann, A., Toloşi, L., Sander, O. and Lengauer, T. (2010), 'Permutation importance: a corrected feature importance measure', *Bioinformatics* **26**(10), 1340–1347.
URL: *https://doi.org/10.1093/bioinformatics/btq134*

Babyak, M. (2004), 'What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models', *Psychosomatic medicine* **66**, 411–21.

Basheer, I. and Hajmeer, M. (2000), 'Artificial neural networks: fundamentals, computing, design, and application', *Journal of Microbiological Methods* **43**(1), 3 – 31. Neural Computing in Microbiology.
URL: *http://www.sciencedirect.com/science/article/pii/S0167701200002013*

Bo-cai, G. (1996), 'Ndwi, a normalized difference water index for remote sensing of vegetation liquid water from space', *Remote Sensing of Environment* **58**(3), 257 – 266.
URL: *http://www.sciencedirect.com/science/article/pii/S0034425796000673*

Breiman, I. (2001), 'Random forests.', *Machine Learning* **45**(1), 5–32.

Brochu, E., Cora, V. and Freitas, N. (2010), 'A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning', *CoRR* **abs/1012.2599**.

Cairns, M., Brown, S., Helmer, E. and Baumgardner, G. (1997), 'Root biomass allocation in the world's upland forests', *Oecologia* **111**, 1–11.

Chen, T. and Guestrin, C. (2016), Xgboost: A scalable tree boosting system, pp. 785–794.

Coulibaly, L., Migolet, P., Adegbidi, H. G., Fournier, R. and Hervet, E. (2008), Mapping aboveground forest biomass from ikonos satellite image and multi-source geospatial data using neural networks and a kriging interpolation, *in* 'IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium', Vol. 3, pp. III – 298–III – 301.

Dallas, C. (2017), 'The "black box" metaphor in machine learning', *Towards Data Science* .
URL: *https://towardsdatascience.com/the-black-box-metaphor-in-machine-learning-4e57a3a1d2b0*

Delegido, J., Alonso, L. and Moreno, J. (2011), 'Evaluation of sentinel-2 red-edge bands for empirical estimation of green lai and chlorophyll content', *Sensors (Basel, Switzerland)* **11**, 7063–81.

EEA (2017), 'European environmental agency. climate change, impacts and vulnerability in europe 2016, an indicator-based report'.

Efroymson, M. (1960), 'Multiple regression analysise', *Ralston A, Wilf HS, editors, Mathematical Methods for Digital Computers* .

FAO (2012), Global ecological zones for fao forest reporting: 2010 update, Technical report, Rome.

Fonti, V. (2017), Feature selection using lasso, Master's thesis, VU Amsterdam.

Forkuor, G., Zoungrana, B., Dimobe, K., Ouattara, B., Vadrevu, K. and Tondoh, E. (2020), 'Above-ground biomass mapping in west african dryland forest using sentinel-1 and 2 datasets - a case study', *Remote Sensing of Environment* **236**, 111496.

Friedman, J. H. (2002), 'Stochastic gradient boosting', *Computational Statistics Data Analysis* **38**(4), 367 – 378. Nonlinear Methods and Data Mining.
URL: *http://www.sciencedirect.com/science/article/pii/S0167947301000652*

Galidaki, G., Zianis, D., Gitas, I., Radoglou, K., Karathanassi, V., Tsakiri–Strati, M., Woodhouse, I. and Mallinis, G. (2017), 'Vegetation biomass estimation with remote sensing: focus on forest and other wooded land over the mediterranean ecosystem', *International Journal of Remote Sensing* **38**(7), 1940–1966.
URL: *https://doi.org/10.1080/01431161.2016.1266113*

Gao, Y., Lu, D., Li, G., Wang, G., Chen, Q., Liu, L. and Li, D. (2018), 'Comparative analysis of modeling algorithms for forest aboveground biomass estimation in a subtropical region', *Remote. Sens.* **10**, 627.

Genuer, R., Poggi, J. M. and Tuleau-Malot, C. (2010), 'Variable selection using random forests', *Pattern Recognition Letters* **31(14)**, 2225–2236.

Gitelson, A. A., Gritz †, Y. and Merzlyak, M. N. (2003), 'Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves', *Journal of Plant Physiology* **160**(3), 271 – 282.
URL: *http://www.sciencedirect.com/science/article/pii/S0176161704704034*

Gitelson, A. A., Kaufman, Y. J. and Merzlyak", M. N. (1996), 'Use of a green channel in remote sensing of global vegetation from eos-modis', *Remote Sensing of Environment* **58**(3), 289 – 298.
URL: *http://www.sciencedirect.com/science/article/pii/S0034425796000727*

Gregorio, M., Ricardo, R. and Marta, M. (2005), 'Producción de biomasa y fijación de co2 por los bosques españoles'.

Guo, Q., Wu, W., Massart, D., Boucon, C. and de Jong, S. (2002), 'Feature selection in principal component analysis of analytical data', *Chemometrics and Intelligent Laboratory Systems* **61**(1), 123 – 132.
URL: *http://www.sciencedirect.com/science/article/pii/S0169743901002039*

Hall-Beyer, M. (2017*a*), 'Glcm texture: A tutorial v. 3.0 march 2017'.

Hall-Beyer, M. (2017*b*), 'Practical guidelines for choosing glcm textures to use in landscape classification tasks over a range of moderate spatial scales', *International Journal of Remote Sensing* **38**(5), 1312–1338.
URL: *https://doi.org/10.1080/01431161.2016.1278314*

Haralick, R., Shanmugam, K. and Dinstein, I. (1973), 'Textural features for image classification', *IEEE Trans Syst Man Cybern* **SMC-3**, 610–621.

Hernandez-Stefanoni, J. L., Castillo-Santiago, M., Mas, J., Wheeler, C., Andres Mauricio, J., Tun-Dzul, F., George, S., Reyes-Palomeque, G., Castellanos-Basto, B., Vaca, R. and Dupuy, J. (2020), 'Improving aboveground biomass maps of tropical dry forests by integrating lidar,alos palsar, climate and field data', *Carbon Balance and Management* **15**.

Hong Han, Xiaoling Guo and Hua Yu (2016), Variable selection using mean decrease accuracy and mean decrease gini based on random forest, *in* '2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)', pp. 219–224.

Huete, A. (1988), 'A soil-adjusted vegetation index (savi)', *Remote Sensing of Environment* **25**(3), 295 – 309.
URL: *http://www.sciencedirect.com/science/article/pii/003442578890106X*

Huete, A. (1994), 'A modified soil adjusted vegetation index', *Remote Sensing of Environment* **48**(2), 119 – 126.

IPCC (2006), '(intergovernmental panel climate change), guidelines for national greenhouse gas inventories volume 4 agriculture,forestry and other land use', **4**, 48.

Kankare, V., Vastaranta, M., Holopainen, M., Räty, M., Yu, X., Hyyppä, J., Hyyppä, H., Alho, P. and Viitala, R. (2013), 'Retrieval of forest aboveground biomass and stem volume with airborne scanning lidar', *Remote Sensing* **5**, 2257–2274.

Kelsey, K. and Neff, J. (2014), 'Estimates of aboveground biomass from texture analysis of landsat imagery', *Remote Sensing* **6**, 6407–6422.

Khan, K., Iqbal, J., Ali, A. and Khan, S. N. (2020), 'Assessment of sentinel-2-derived vegetation indices for the estimation of above-ground biomass/carbon stock, temporal deforestation and carbon emissions estimation in the moist temperate forests of pakistan', *Applied Ecology and Environmental Research* **18**, 783–815.

Kvalseth, T. O. (1983), 'Note on the r2 measure of goodness of fit for nonlinear models', *Bulletin of the Psychonomic Society* **21(I)**, 79–80.

Lackner, K. S. (2003), 'A guide to co2 sequestration', *Science* **300**(5626), 1677–1678.
URL: *https://science.sciencemag.org/content/300/5626/1677*

Lewinson, E. (2019), 'Explaining feature importance by example of a random forest', *Towards Data Science* .
URL: *https://towardsdatascience.com/explaining-feature-importance-by-example-of-a-random-forest-d9166011959e*

Liu, Y., Guirui, Y., Wang, Q.-F. and Zhang, Y. (2012), 'Huge carbon sequestration potential in global forests', *Journal of Resources and Ecology* **3**, 193–201.

Lu, D. (2006), 'The potential and challenge of remote sensing-based biomass estimation', *International Journal of Remote Sensing* **27**(7), 1297–1328.
URL: *https://doi.org/10.1080/01431160500486732*

Lu, D., Chen, Q., Wang, G., Liu, L., Li, G. and Moran, E. (2016), 'A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems', *International Journal of Digital Earth* **9**(1), 63–105.
URL: *https://doi.org/10.1080/17538947.2014.990526*

Lundberg, S., Erion, G., Chen, H., DeGrave, A., Prutkin, J., Nair, B., Katz, R., Himmelfarb, J., Bansal, N. and Lee, S.-I. (2020), 'From local explanations to global understanding with explainable ai for trees', *Nature Machine Intelligence* **2**.

Lundberg, S. and Lee, S.-I. (2017), A unified approach to interpreting model predictions, *in* 'NIPS'.

Malhi, A. and Gao, R. X. (2004), 'Pca-based feature selection scheme for machine defect classification', *IEEE Transactions on Instrumentation and Measurement* **53**(6), 1517–1525.

Muthukrishnan, R. and Rohini, R. (2016), Lasso: A feature selection technique in predictive modeling for machine learning, pp. 18–20.

Nogueira, F. (2014), 'Bayesian Optimization: Open source constrained global optimization tool for Python'.
URL: *https://github.com/fmfn/BayesianOptimization*

Nuthammachot, N., Askar, A., Stratoulias, D. and Wicaksono, P. (2020), 'Combined use of sentinel-1 and sentinel-2 data for improving above-ground biomass estimation', *Geocarto International* **0**(0), 1–11.
URL: *https://doi.org/10.1080/10106049.2020.1726507*

Pandit, S., Tsuyuki, S. and Dube, T. (2018), 'Estimating above-ground biomass in subtropical buffer zone community forests, nepal, using sentinel 2 data', *Remote. Sens.* **10**, 601.

Pandit, S., Tsuyuki, S. and Dube, T. (2019), 'Exploring the inclusion of sentinel-2 msi texture metrics in above-ground biomass estimation in the community forest of nepal'.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011), 'Scikit-learn: Machine learning in Python', *Journal of Machine Learning Research* **12**, 2825–2830.

Picard, N., Saint-André, L. and Henry, M. (2012), Manual for building tree volume and biomass allometric equations: from field measurement to prediction.

Pinty, B. and Verstraete, M. (1992), 'Gemi: A non-linear index to monitor global vegetation from satellites', *Vegetatio* **101**, 15–20.

Puliti, S., Hauglin, M., Breidenbach, J., Montesano, P., Neigh, C., Rahlf, J., Solberg, S., Klingenberg, T. F. and Astrup, R. (2020), 'Modelling above-ground biomass stock over norway using national forest inventory data with arcticdem and sentinel-2 data', *Remote Sensing of Environment* **236**.

Rouse, J., Haas, R. H., Schell, J. A. and Deering, D. (1973), Monitoring vegetation systems in the great plains with erts.

Salem, I., Basam, D., Taoufik, K. and Nazmi, S. (2020), 'A review of terrestrial carbon assessment methods using geo-spatial technologies with emphasis on arid lands', *Remote Sensing* **12**(12), 2008.
URL: *http://dx.doi.org/10.3390/rs12122008*

Schapire, R. (1990), 'The strength of weak learnability', *Machine Learning* **5**, 197–227.

Schimel, D., House, J., Hibbard, K., Bousquet, P., Ciais, P., Peylin, P., Braswell, B., Apps, M., Baker, D., Bondeau, A., Canadell, J., Churkina, G., Cramer, W., Denning, A., Field, C., Friedlingstein, P., Goodale, C., Heimann, M., Houghton, R., Melillo, J., Moore, B., Murdiyarso, D., Noble, M., Pacala, S., Prentice, I., Raupach, M., Rayner, P., Scholes, R., Steffen, W. and Wirth, C. (2001), 'Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems', *Nature* **414**(6860), 169–172. Publisher: Nature Publishing Group.

Seif, G. (2019), 'A beginner's guide to xgboost', *Towards Data Science* .
URL: *https://towardsdatascience.com/a-beginners-guide-to-xgboost-87f5d4c30ed7*

Sharma, S. (2017), 'Epoch vs batch size vs iterations', *Towards Data Science* .
URL: *https://towardsdatascience.com/epoch-vs-iterations-vs-batch-size-4dfb9c7ce9c9*

Smith, G. (2018), 'Step away from stepwise', *Journal of Big Data* **5**, 32.

Snoek, J., Larochelle, H. and Adams, R. (2012), 'Practical bayesian optimization of machine learning algorithms', *Advances in Neural Information Processing Systems* **4**.

Spiess, A.-N. and Neumeyer, N. (2010), 'An evaluation of r2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: A monte carlo approach', *BMC pharmacology* **10**, 6.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014), 'Dropout: A simple way to prevent neural networks from overfitting', *Journal of Machine Learning Research* **15**, 1929–1958.

Stinson, G., Kurz, W. A., Smyth, C. E., Neilson, E. T., DymondY, C. C., Metsaranta, J. M., Boisvenue, C., Rampley, G. J., Li, Q., White, T. M. and Blain, D. (2011), 'An inventory-based analysis of canada's managed forest carbon dynamics, 1990 to 2008', *Global Change Biology* **17**(6), 2227–2244.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2486.2010.02369.x*

Sutton, C. D. (2005), 11 - classification and regression trees, bagging, and boosting, *in* C. Rao, E. Wegman and J. Solka, eds, 'Data Mining and Data Visualization', Vol. 24 of *Handbook of Statistics*, Elsevier, pp. 303 – 329.
URL: *http://www.sciencedirect.com/science/article/pii/S0169716104240111*

Thomas, S. and Martin, A. R. (2012), 'Carbon content of tree tissues: A synthesis', *Forests* **3**, 332–352.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society* **58(1)**, 267–288.

Torralba and Crespo-Peremarch (2018), Assessing the use of discrete , full-waveform lidar and tls to classify mediterranean forest species composition.

UNFCCC (1997), 'Kyoto protocol to the united nations framework convention on climate change'.

Vashum, K. (2012), 'Methods to estimate above-ground biomass and carbon stock in natural forests - a review', *Journal of Ecosystem Ecography* **02**.

Vicharnakorn, P., Shrestha, R., Nagai, M., Salam, A. P. and Kiratiprayoon, S. (2014), 'Carbon stock assessment using remote sensing and forest inventory data in savannakhet, lao pdr', *Remote. Sens.* **6**, 5452–5479.

Walker, B. (2019), 'Pca is not feature selection. what it actually does and when you can and can't use it.', *Towards Data Science* .
URL: *https://towardsdatascience.com/pca-is-not-feature-selection-3344fb764ae6*

Wani, A., Joshi, P. K. and Singh, O. (2015), 'Estimating biomass and carbon mitigation of temperate coniferous forests using spectral modeling and field inventory data', *Ecol. Informatics* **25**, 63–70.

Zhao, P., Lu, D., Wang, G., Wu, C., Huang, Y. and Yu, S. (2016), 'Examining spectral reflectance saturation in landsat imagery and corresponding solutions to improve forest aboveground biomass estimation', *Remote Sensing* **8**.

Zheng, D., Rademacher, J., Chen, J., Crow, T., Bresee, M. K., Moine, J. L. and Ryu, S. R. (2004), 'Estimating aboveground biomass using landsat 7 etm+ data across a managed landscape in northern wisconsin, usa', *Remote Sensing of Environment* **93**, 402–411.

Appendix A

Coefficient of determination (R2)

The following Appendix contains a brief introduction on the coefficient of determination (R2) and shortly covers the reason why this is not an adequate measure for non-linear models. Lastly, this section includes an explanation on why it is safe to assume that the problem object of this study can be considered non-linear.

The inadequacy of using R2 as a measure of goodness of fit for non-linear models has been well known within the mathematical literature for many years (Kvalseth, 1983). Nevertheless, R2 keeps often being used by the scientific community in many fields of study such as vegetation mapping. Indeed, this was the case for all of the studies on Above Ground Biomass (AGB) estimation that were reviewed in this work.

R2 is a metric of correlation easy to compute and straightforward to interpret. Its value is given by:

$$R^2 = \frac{Explained\ variance}{Total\ variance} \tag{6.1}$$

Where:

$$Total\ variance = Explained\ variance\ + Error\ variance \tag{6.2}$$

This arrangement ends up in R2 ranging from 0 to 1. Such an assumption holds true in linear models, however, this is not the case in non-linear models because the sum between the variance explained by the model and the error variance does not correspond to the total variance (Kvalseth, 1983; Spiess and Neumeyer, 2010). Therefore, the underlying assumptions for R2 are invalidated when this measure is used for non-linear models.

In a Monte Carlo simulation carried out by Spiess and Neumeyer (2010), the performance of R2 for validating non-linear models was tested. After performing thousands of simulations, they found out that using R2 for assessing the goodness of fit in non-linear models leads to false conclusions about the best performing models. Hence, R2 was found to be high for both effective and terrible models, and it would not always increase in value whenever better performing models were presented. They conclude that using R2 for choosing the most appropriate regression model amongst non-linear ones rarely leads to

choosing the actual best model. Hence, R2 was not able to discriminate between good and bad non-linear models.

This thesis tested a palette of Machine Learning algorithms for the estimation of Above Ground Biomass (AGB). Thus, with the purpose of proving the non-linearity of the problem investigated, the relationships between each of the explanatory variables and the target were tested. This was achieved by fitting a linear regression and a quadratic polynomial for each of the 63 features and the measured AGB (target); therefore, it was assessed whether a higher amount of variance was explained by the linear or by the quadratic model. Hence, figure 6.1 illustrates how the two models were fitted to predict AGB by using one of the features - the Normalized Difference Vegetation Index from the Summer image. Both the feature and the target were normalized with the scaler explained in the main text, Chapter 3.

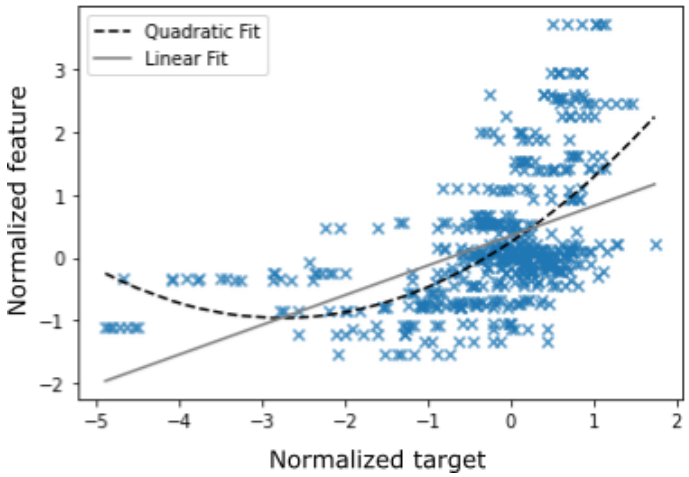


Figure 6.1: Testing of a linear and a quadratic model.

Subsequently, the ratio between the amount of variance explained by the linear model and the one explained by the quadratic model (Equation 6.3) was represented in figure 6.2. Thus, for any feature for which the value of this ratio is not equal to or larger than one, the relationship between that feature and the target cannot be considered to be linear. In fact, by fitting a linear model on non-linear data, the variance explained by that model is going to be lower than if using a non-linear model; hence, using a line only explains the amount of variance over one direction. Therefore, figure 6.2 shows that, every features, the relationship between them and the target is better represented by the non-linear model. Additionally, it must be kept in mind that this simple experiment only took into account one feature at the time, therefore it does not consider the possible non-linear interaction between features whenever a model is built using more than one, as it is often the case.

$$Explained\ variance = \frac{Explained\ variance\ (linear)}{Explained\ variance\ (quadratic)} \tag{6.3}$$

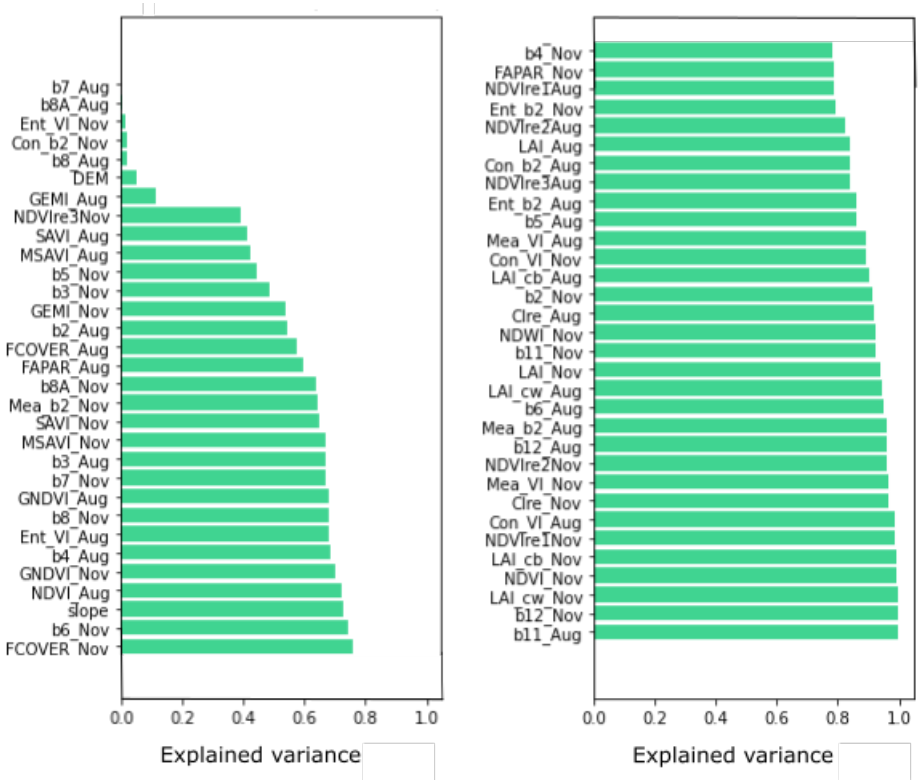


Figure 6.2: Ratio of explained variance of a linear and a quadratic model.

Masters Program in **Geospatial Technologies**

Supported by:
Education and Culture
ERASMUS MUNDUS