Masters Program in Geospatial Technologies

# *THE USE OF REMOTELY SENSED DATA FOR FOREST BIOMASS MONITORING*

## *A case of forest sites in North-Eastern Armenia*

Manvel Khudinyan

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

# THE USE OF REMOTELY SENSED DATA FOR FOREST BIOMASS MONITORING

## A case of forest sites in North-Eastern Armenia

Dissertation supervised by:

**Joel Silva, PhD.** Information Management School, Nova University of Lisbon, Lisbon, Portugal

**Gohar Ghazaryan, MSc.** Center for Remote Sensing of Land Surfaces (ZFL), University of Bonn, Bonn, Germany

**Ignacio Guerrero, PhD.** Universitat Jaume I, Castellón de la Plana, Spain

February, 2019

ACKNOWLEDGMENTS

I am especially indebted to my supervisor PhD Joel Silva for the great support and technical guidance he has offered. I would also like to particularly thank my co-supervisor PhD Gohar Ghazaryan for her continuous support and constructive comments. I am also very thankful to co-supervisor Prof. Ignacio for the time he has devoted to my thesis.

This work would not have been possible without Prof. Dr. Marco Painho, Prof. Brox and every member of staff at UNL and ifgi who I have to thank for my wonderful study experience.

I would especially like to thank Prof. Hovik Sayadyan who provided the field data (no data – no thesis), for supporting my ideas and for the encouragement. I am also grateful to PhD Gorik Avetisyan and Arman Kandaryan for the data they provided, as well as my friends, foresters PhD Vahe Martirosyan and PhD Vahe Matsakyan for the consultations on the forests and their help.

Last but not least I would like to thank my family, my friends and my classmates for their constant support and care.

# THE USE OF REMOTELY SENSED DATA FOR FOREST BIOMASS MONITORING

## A case of forest sites in North-Eastern Armenia


## ABSTRACT

In recent years there has been an increasing interest in the use of synthetic aperture radar (SAR) data and geospatial technologies for environmental monitoring. Particularly, forest biomass evaluation was of high importance, as forests have a crucial role in global carbon emission. Within this study we evaluate the use of Sentinel 1 C-band multitemporal SAR data with combination of Alos Palsar L-band SAR and Sentinel 2 multispectral remote sensing (RS) data for mapping forest aboveground biomass (AGB) of dry subtropical forests in mountainous areas. Field observation from National Forest Inventory was used as a ground truth data. As the SAR data suffers greatly by the complex topography, a simple approach of aspect and slope information as forestry ancillary data was implemented directly in the regression model for the first time to mitigate the topography effect on radar backscattering value. Dense time-series analysis allowed us to overcome the SAR saturation by the forest phenology and select the optimal C-band scene. Image texture measures of SAR data has been strongly related to the biomass distribution and has robustly contributed to the prediction. Multilinear Stepwise Regression allowed to select and evaluate the most relevant variables for AGB. The prediction model combining RS with ancillary data explained the 62 % of variance with root-mean-square error of 56.6 t $ha^{-1}$. The study also reveals that C-band SAR data on forest biomass prediction is limited due to their short wavelength. Further, the mountainous condition is a major constraint for AGB estimation. Additionally, this research demonstrates a positive outcome in forest AGB prediction with freely accessible RS data.

# KEYWORDS

Synthetic Aperture Radar

Forest

Aboveground Forest Biomass

Multilinear Stepwise regression

Sentinel 1

Sentinel 2

ALOS PALSAR

GLCM

# ACRONYMS

**AGB** – Aboveground Biomass

**SAR –** Synthetic Aperture Radar

**REDD+ –** Reducing Emission from Deforestation and forest Degradation

**UN –** European Union

**NDVI** – Normalized Difference Vegetation Index

**NDMI** – Normalized Difference Moisture Index

**RVI** – Simple Ratio Vegetation Index

**FAO** – Food and Agriculture Organization of the United Nations

**NFI** – National Forest Inventory

**ESA** – European Space Agency

**SRTM -** Shuttle Radar Topography Mission

**SWR** – Step Wise Regression

**JAXA** – Japanese Aerospace Exploration Agency

**DBHOB** – Diameter on Brest Height Over Bark

**GEE** – Google Earth Engie

**UNDP –** United Nation Development Program

**GEF –** Global Environmental Facility

INDEX OF THE TEXT

Page

## INDEX OF TABLES

## INDEX OF FIGURES

# INTRODUCTION

## 1.1 Theoretical Framework

Forests cover around 30% of the land surface (Zhang et al., 2018) and have the highest capability for carbon segregation and storing capacity compared to any other terrestrial ecosystem (about 80% of terrestrial biosphere carbon storage) (Chen et al., 2018). Thus they are of paramount importance in mitigation of global climate change (Zhang et al., 2018). Forest ecosystems are identified by the Inter-Governmental Panel on Climate Change through the Kyoto Protocol, as one of the main so-called "brakes" against the global climate change problems (Gibbs et al., 2007, Timothy et al., 2016). Once the forest is cut or degraded, the stored carbon is discharged onto the atmosphere in a form of carbon dioxide (CO2) (Gibbs et al., 2007). Hence, the estimation of forest carbon stock is needed for reducing the uncertainty in the global carbon budget, also, for forest management and planning and biodiversity conservation purposes (Laurin et al., 2016).

Nowadays, in terms of global ecological crisis and forest degradation caused for carbon stock reduction all over the world, UN-REDD+ (Reducing Emissions from Deforestation and forest Degradation) assumes for participant developing countries to have a carbon monitoring system (Kachamba et al., 2016). Forest aboveground biomass (AGB), its amount and spatial distributions are the central components for evaluating forest carbon budgets (Soenen et al., 2010). Therefore, the quantification, mapping and monitoring of forest AGB, which consist of tree and shrub aggregated Net Primary Productivity, has become a broad topic in international climate change conferences (Timothy et al., 2016, Kumar, 2015).

Until now, the best results for AGB estimation are captured through national forest inventories (NFIs) using experimental site establishment and destructive forest cut or tree allometric equations and statistical methods (Soenen et al., 2010). Then, this data is subsequently being used to estimate country level biomass and carbon stock estimates. However, these conventional methods are challenging to enhance and upscale over large areas as they are limited to where inventory data are measured, are spatially incomplete and commonly sparse. Also, another drawback of these methods is the limited frequency of field data collection campaigns and existing information

are often outdated, as they are usually done one time only, thus, they are not reliable source for future carbon budget reporting. A detailed forest inventory takes a substantial amount of time, leads to significant costs, and does not permit a synoptic view of the distribution of biomass across a forest landscape (Goetz & Dubayah, 2011; Kachamba et al., 2016; Santoro & Cartus, 2018).

Nonetheless, many developing countries, including Armenia, do not have exhaustive NFIs. Armenia is currently in the ongoing stage of implementing the REDD+ pilot mechanism in north-eastern (NE) forests of the country and NFI is one of the activities that is being covered already. The REDD+ concept is to provide financial encouragement to help the developing countries voluntarily act against the deforestation rate and consequently decrease the carbon emission and keep it in balance. Those countries that can succeed with emission reduction can then sell their carbon credits in the international carbon market. The aim of the abovementioned projects is to stimulate global emission reduction and to contribute the conservation of biodiversity and protect other ecosystems (Gibbs et al., 2007).

A promising approach aimed at lessening labor and operational costs, at the same time improving the reliability of estimated biomass gained by field sampling process, combines data from ground-based forest inventories and remotely sensed data. Last development of remote sensing technologies and successive methodological innovations enable researchers for rapid assessment of appropriate data on a large area at a lower cost (Kachamba 2016). In these terms it becomes more important to explore and develop prediction models using satellite imagery to contribute to the estimation of aboveground biomass.

Remotely sensed data for forest monitoring applications are divided into three primary sources and systems: optical imagery (both satellite and aerial), Airborne Laser Scanning (ALS), and RADAR (e.g., synthetic aperture radar (SAR)). All of them have been widely applied for forestry purposes in majority of the countries with different successes rates depending on the data type, forest structure, geographic features of the implied area and on other factors. Optical sensors were mainly applied for horizontal forest structure generation and for the AGB assessment through field sampling, based on their spectral signatures (reflectance and vegetation indices). The advantages of optical sensing can be the global cover, frequency of repetitiveness and cost suitability. The limitations of optical sensors are the poor penetration capability for

clouds and forest canopies, also, they are affected by data saturation. Most used optical platforms are IKONOS, Quickbird, Worldview, ZY-3, SPOT, Sentinel, Landsat (Chen et al., 2018; Gibbs et al., 2007; Kachamba et al., 2016).

The SAR systems show capabilities of improving the accuracy of forest AGB acquisition as the ability of canopies penetration is much higher than the one in optical, thus, SAR imagery can give information on vertical structure and volumetric features of the forest. In addition, SAR data are not affected by weather and light conditions. However, the radar scanning has limitations such as signal saturation in complex structured forests and the poor ability of forest type distinguishing (Gibbs et al., 2007). The most used sensors for forest ABG retrieval are Palsar, Palsar 2, Sentinel 1, TerraSAR-X (Berninger et al., 2018; Chen et al., 2018; Laurin et al., 2018).

Thus, SAR becomes one of the most useful tools to gain vertical structure and volumetric features of the forest (Laurin 2018). Sentinel 1 C-band SAR multitemporal images, that are available since 2014 and have high temporal recurrence and fine spatial resolution, have become very useful tool for forest AGB estimation and forest monitoring in the last few years (Joshi et al., 2016). Literature review shows many cases of forest biomass estimation based on Sentinel 1 data. In addition, better results are recorded when different SAR data (L-band and C-band) is conducted with optical data (Laurin et al., 2016). Establishing the balance between the proper use of natural resources and current environmental legislation is one of the functions that the environmental institutions play with obvious increase of the complexity degree, citing the need for automation of analysis processes that help the decision-making, in order to utilize the available resources in a more efficient manner.

## 1.2 Statement of the Problem

The forest degradation is an ongoing problem in Armenia and it suffers mostly by illegal and not well-controlled forest logging and over-grazing. In general, the forest borders are not shrinking and the forest cover area does not transform significantly (Nayemberyan forest enterprise increased the area of 296 ha after the last forest boundary update) (FMP-Noyemberyan, 2018). Even thoughthe quality of the forests and the amount of biomass in forests change substantially (Sayadyan et al., 2018).

Notably, a massive deforestation occurred from 1990-1997 during the energetic crises, when the fuelwood was the primary mean of house heating for the majority of the households, especially, in the settlements close to the forest. Nowadays, the felling process is decreased and stabilized, but still forest monitoring center has proven that 95% of logging is no coincide with the FMPs. Also, some estimations show that around 630,000 $m^3$ of wood is still illegally felled in Armenia annually (UNDP-GEF, 2015). The forest cover in Armenia is highly fragmented and the biggest part (approximately 62%) with the most valuable forests are located in the NE of Armenia, namely in the Lori and Tavush provinces (territorial admin called "marz".). Those provinces are very poor (poverty rate: 38.6% and 27.7% respectively) and the population is still highly dependent on the forest supply and, subsequently, the pressure on the forest still remains significantly high (FMP-Nayemberyan, 2018; UNDP-GEF, 2015).

The economic reasons, accompanied by other factors such as the lack of the professionalism of the authorities, inadequate planning, the lack of interest of the local population in forest management and conservation, the lack of financial recourses for respective bodies are seen as the main driver factors of deforestation (to organize better forest management, conservation and monitoring). These reasons lead to unsustainable forest management with an output of deforestation and forest degradation, lots of valuable biodiversity and carbon sequestration potential reduction (Sayadyan et al., 2018; UNDP-GEF, 2015).

Currently, the Armenian government together with UNDP and GEF and with a support of other donor organizations are implementing sustainable forest management plan by aligning those with REDD+ and GEF objectives in the NE forests of Armenia. According to this project, one of the key conditions of the forest management is the effective and accurate estimation and monitoring over the forest resources, particularly, forest biomass, which has the main carbon dioxide absorption capacity (Moreno-Sanchez et al., 2007, UNDP-GEF, 2015).

So far, the use of earth observation derived datasets for forest inventory and monitoring in Armenia is not fully implemented. In the scope of the abovementioned project there were Geographic Information Systems and remote sensing methods implemented but as tools for accurate forest demarcation. There were no studies found about the usage of remotely sensed data in forest biomass estimation and forest

resource monitoring in Armenia. Although, the use of multisource remote sensing data for forest biomass and carbon stockis are discussed in several studies. The overwhelming majority of those studies were conducted for tropical, subtropical and boreal forests, which contain the most part of the global forest cover. Even if the NFI project in Armenia is being implemented as a cost effective NFI, assuming that the field sampling net is low intensity and sampling plot size is decreased in area, the project is still costly and not sensible to use later on for forest monitoring and data updating purposes (Haywood et al., 2018).

Thus, this research is aimed at implementing remote sensing methods for forest monitoring, especially if the integration of RS data can substitute field observations.

## 1.3 Aims and Objectives

The aim of this study lies in suitability analysis of C-band and L-band SAR and HR optical imagery for AGB estimation and the integration with ancillary data in dry subtropical mixed forests in the mountainous regions.

- To study the sensitivity of C-band and L-band backscatter data to forests biomass.
- To study the dependence of C-band SAR data on the forest phenology.
- To apply multisource satellite data combination for forest biomass monitoring.

To fulfill the aims and specific objectives, the following research questions need to be answered:

- Can multisource remotely sensed data be integrated in optimizing the forest monitoring?
- Is Sentinel 1 C-band backscatter dependent on forest seasonality? Are the dense time series able to handle the backscatter saturation problem?
- Are the image texture measures relevant to forest biomass.
- Is it possible to robustly derive remote sensing estimation of AGB using regression models?

## 1.4 Outline

The structure of the dissertation is as follows: **Chapter 2** presents the review of existing literature on the topic of this research. **Chapter 3** describes the conditions and specification of the study area and the data used within this study. **Chapter 4** discusses the approaches and methodology, also describes the data pre-processing steps in very details. **Chapter 5** presents and discusses the results for forest/non-forest classification and for regression analysis, also remarks the limitations and future possible researches. Section **Conclusion** sums up the results of this research and evaluates its contributions.

# LITERATURE REVIEW

## 2.1 Introduction for Aboveground Forest Biomass

Before starting the analysis of remotely sensed data for forest aboveground biomass estimation, it is essential to understand the concept of forest AGB, the traditional techniques and methods of estimation, which gives better understanding and vision of the parameters and conditions should be considered before applying remote sensing methods.

Aboveground live biomass includes the living biomass of trees, shrubs and herbs above the soil including stem, stump, branches, bark, leaves (UNDP-GEF 00091048, 2015), and represents the largest pool of carbon stock (Gibbs et al., 2007). AGB is widely used for correlations and estimations of carbon storage in some of the other pools such as root biomass and consequently the carbon stocks (approximately 20 % of AGB (Gibbs et al., 2007)) and dead wood or litter carbon pool (appr. 10-20 % in natural forests (UNDP-GEF 00091048, 2015)). Thus, the forest AGB estimation is an essential step in quantifying the carbon stocks in the forests.

There are several methods developed for AGB estimation with different demands and level of accuracy. The most accurate and straightforward method to quantify the forest biomass is establishing sampling sites on the field and harvesting all the trees, drying them (in particular, for carbon stock estimation) and weighting the biomass (Gibbs et al., 2007). This method is very expensive and has destructive affection on the experimental site. It is precise for that one location only, and impractical in other regions for country level analysis (Chen et al., 2018).

Such methods cannot be applied across the landscape. Consequently, many investments were put into development of models that are able to “scale up” the field measurement results over larger areas. Gibbs did sum up in glance the main methods used for forest AGB estimation with their advantages and limitations (**Table 1**) (Gibbs et al., 2007).

| Method | Benefits | Limitations | Uncertainty |
|---|---|---|---|
| Biome averages | • Immediately available could increase accuracy • Data refinements could increase accuracy • Globally consistent | • Fairly generalized • Data sources not properly sampled to describe large areas | High |
| Forest inventory | • Generic relationships readily available • Low-tech method widely understood • Can be relatively expensive as field-labor is largest cost | • Generic relationships widely understood • Low-tech method inexpensive as • Can be relatively field-labor is largest cost | Low |
| Optical remote sensors | • Satellite data routinely available at global scale • Globally consistent | • Limited ability to develop good models for tropical • Spectral indices saturate at relatively low carbon stocks • Can be technically demanding | High |
| Radar remote sensors | • Satellite data are not always free •New systems launched are expected to provide improved data • accurate for young or sparse forest | • Less accurate in mature forests because signal saturates also increases errors • Mountainous terrain also increases errors • Can be expensive | Medium |

**Table 1.** Benefits and limitations of available methods to estimate national-level forest carbon stocks (data source: Gibbs et al., 2007)

Bouvet and Gibbs are dividing biome-average method into two main steps (Bouvet et al., 2018; Gibbs et al., 2007). First, data gathering is being done through NFIs. Until now the best results are gained by field calculations using experimental plots, harvesting and actual forest volume estimation (Kachamba et al., 2016). In second step allometric models are applied to calculate average biomass for a sampling plot located within a certain stratum. Later, national level of carbon stock can be predicted by applying the average biomass and carbon density values (de Badts, 2002) over the region for the same forest strata (Bouvet et al., 2018; Kachamba et al., 2016). Unlike the high level of uncertainty compared to the forest harvesting method, this method can be immediately available and provide information about AGB on a wide scale landscape.

Ground-based forest inventory focused on field campaigns and forest inventory measurements for estimation of forest AGB. Measurements include diameter at breast height (DBH), average tree height and tree type, and through which forest biomass is being calculated using allometric relationships (Chave et al., 2005; Gibbs et al., 2007). Only DBH is describing AGB with 95 % accuracy (Chave et al., 2005). Zhang offers to collect data about the vegetation type and soil type as well (Zhang et al., 2018). As the AGB consist of stem, bark, branches and leaves or needles depending on the forest

type, he assumes to organize field data acquisition differently for coniferous and deciduous forests and exclude leave biomass for winter calculations in deciduous forests (Clerici et al., 2016).

For field data collection there are several approaches for sampling points selection. Systematic and random sampling designs are the most used ones. Systematic sampling uses? regular grid for plot location selection, and the random selection is applied to randomly allocated sampling points. Both of the approaches are not considering forest type stratification and forest distribution. This, both schemes can under- or over-sample as far as the patterns in nature have naturally clumpy and random distribution is less likely (Gibbs et al., 2007).

One of the uncertainties of AGB estimation by converting tree measurements data is the lack of standard method. Literature review shows different approaches by different authors for AGB estimation even for the same type of forests (Chave et al., 2005; Clerici et al., 2016; Gibbs et al., 2007). Gibbs presents a complete overview of forest AGB estimations by different authors for the same countries, and as the article points out, for Shri-Lanka the AGB estimation by different methods can alter in the range of 138–509 Mg/m2. The uncertainty occurs when shifting from one forest type to another. In tropical forests 1 ha forest area can contain 300 different tree species. This means that one cannot use species-specific regression model, which is common to apply while working with temperate or boreal forests. In such conditions better results are achieved by mixed-species regression model (Chave et al., 2005).

Amongst existing number of statistical methods for AGB acquisition based on sampled allometric measurements, Chaves suggests "biome-diameter-height regression" and "biome-diameter regression" models for their simplicity and wide application (Chave et al., 2005). Biome-diameter-height regression includes information about tree height (H), diameter (D), and wood specific gravity (ρ) (Dawakin's regression model):

$$\ln(AGB) = \alpha + \ln(D^2 H \rho) \quad (1)$$

Both for the regression and the rest are applied using linear models.

Field sampling data does not always include the tree height information. In this case Chaves assumes relationship between logarithm of height, In(H), and the logarithm of diameter, In(D). A polynomial relationship between log. height and log. diameter

gives a reasonable generalization of the power-low model. Hence, the model will have following equation:

$$\ln(AGB) = a + b*\ln(D^2H\rho) + c*\ln(D^2) + d*\ln(D^3) + \beta*3\ln(\rho) \quad (2)$$

The power-low is parametrized as c=d=0 (Chave et al., 2005).

All those conventional models are accurate for the location of sampling site but are difficult to extend over large area and are being done not more than once for certain area because of paramount resource demand. Instead, remotely sensed data provides useful for AGB estimation data for the whole globe and with temporal resolution, but cannot measure the biomass directly, thus require ground-truth data (Gibbs et al., 2007; Kumar et al., 2015). Another important advantage of remote sensing is the ability to easily collect data on the places where ground access is limited.

Optical remote sensing data from different sensors (see chapter 1.1) have been implemented (tried) for forest AGB indirect estimation via statistical relationship between field measured data and satellite-observed vegetation indices (Fassnacht et al., 2014). As the optical sensor records the information reflected from the forest canopy and is dependent on the leaf structure, pigmentation and moisture and have week penetration possibilities, thus the uncertainty is high, especially, in dense tropical forests with big amount of biomass (Chen et al., 2018; Joshi et al., 2016; Treuhaft et al., 2017). In the contrary, comparatively low uncertainty and higher correlation between sensor-data and field-data are recorded in boreal forests where the hierarchical structure is absent in the forests and individual trees can be fixed on the photos (Gibbs et al., 2007; Kachamba et al., 2016). Clerici affirms that very high-resolution (VHR) optical imagery improves the AGB estimations in forests (Clerici et al., 2016) using ratio vegetation index (RVI), normalized vegetation index (NDVI), transformed vegetation index (TVI), while the Vegetation Index Number (VIN) recorded the worst performing. Kumar offers a use of hyperspectral sensor data, particularly underlying the importance of mid-infrared (MIR) reflectance as the best describer for forest biomass. He states that MIR has advantages over visible and near-infrared (NIR) reflectance (L. Kumar et al., 2015). Signal saturation in the forests is the main constrain for optical remote sensing for AGB estimation and almost all the reviewed articles detected to be suffering with this issue.

With the launch of different SAR sensors with freely distributed data, there is increasing interest in the use of these observations for forest AGB estimation. This interest started in 2000s when TerraSAR-X, Alos, Palsar L-band were launched, and become more popular, when in 2014 first Sentinel 1 C-band data became freely available. For instance, the majority of the studies for SAR in AGB estimations were published in the last 3-5 years (Berenguer et al., 2018; Berninger et al., 2018; Clerici et al., 2016; Joshi et al., 2016; Kumar et al., 2015; Navarro et al., 2016; Reiche et al., 2018; Santi et al., 2017; Santoro & Cartus, 2018; Vafaei et al., 2018; Yu & Saatchi, 2016).

Generally, the most regularly used methods for forest AGB estimation using SAR data can be classified into several groups. The simplest and subsequently the most used method is the measuring backscattering coefficient from Polarimetric SAR (PolSAR) to reveal the roughness of the surface. This is called 2D PolSAR method (Zhang et al., 2018). As it was mentioned before, the best results are obtained with cross-polarization dual polarization which is more sensitive to the forest AGB. Most of them applied logarithm of biomass and used the backscattering coefficient for forest biomass prediction (Laurin et al., 2018; Zhang et al., 2018). This method still has some limitations such as saturation, which depends on the wavelength and the incidence angel (Joshi et al., 2016). The capabilities, advantages and drawbacks of SAR backscatter data in forest AGB estimation, will be discussed in details in the next subchapters

Interferometric SAR (InSAR) or Polarimetric InSAR (PolInSAR) by using interferometric phase or coherence tomography of InSAR are able to record the elevation of the ground and the top of the forest, and thus, they can derive the forest height which then can be transformed into forest biomass by allometric equation models. This model is promising and has more potential, and it can reduce the saturation in some extant (Chen et al., 2018). A number of studies show improved AGB estimation results by using InSAR data (Huang, Ziniti, Torbick, & Ducey, 2018; Santoro & Cartus, 2018; Zhang et al., 2018).

There are also other methods for AGB estimation through Laser scanning and LiDAR scanning (Joshi et al., 2016), which are not fitting in the scopes of this research, therefore they will not be discussed in details.

## 2.2 SAR and Optical remote sensing data for AGB estimation

SAR are active microwave radars operating in between 1 centimeter and several meters of electromagnetic spectrum. A part of the impulse energy from radar, once meeting the Earth's surface, it is being scattered back towards the sensor. This is what the Radar records and measures. Depending on the surface type and structure, different amount of energy is scattered back and thus, different parameters of the surface are recorded on the image. The image intensities depend on the backscattered signal characteristics such as wavelength, incidence angle, signal polarization, scan direction, in addition to such parameters as surface roughness, moisture, dielectric properties, geometric shape, orientation (Pohl C., 2017). The frequency, incidence angle, polarization and scan direction are properties of system and the dielectric characteristics, orientation, surface roughness and the moisture are target properties (Periasamy et al., 2018). Those are the parameters that are influencing on the backscattered signal and thus, describing the surface objects, they are core for radar image processing.

SAR images for the same area vary as the system parameters change. For instance, the penetration capability of the signal is dependent on the wavelength: The longer wavelength the larger objects they penetrate. According to the rule of thumb the penetration length is the half of the wavelength. According to the same rule, higher the backscatter intensity the rougher the surface that is being imaged (CRISP 2001, n.d.).

| Radar Band | Wavelength (cm) | Frequency (GHz) |
|---|---|---|
| X | 2.5–3.75 | 8–12 |
| C | 3.75–7.5 | 4–8 |
| S | 11.11–7.69 | 2.7–3.9 |
| L | 15–30 | 1–2 |
| P | 100 | 0.3 |

**Table 2.** Common radar bands used in Remote Sensing (data source: Pohl. 2017)

Depending on the used wavelength, SAR sensors are divided into several bands (namely: K, X, S, C, L, P, VHF) (Timothy et al., 2016). **Table 2** describes the wavelength and appropriate frequencies for the SAR bands most commonly used for land cover surveys.

L and P- band (15-100 cm together) have privileges over C-band (3.75-7.5 cm) in terms of forest vertical values description. C-band can penetrate only leaves and

needles and small branches, while L and P-bands provide stronger backscatter form trunk and large branches (**Figure 1**) (Joshi et al., 2016). This leads to a different result when calculating forest AGB trough above-mentioned SAR bands. According to the reviewed literature L-band analysis records higher correlation with ground data compared with the C-band (Bouvet et al., 2018; Laurin et al., 2016, 2018; Mermoz et al., 2015; Yu & Saatchi, 2016).

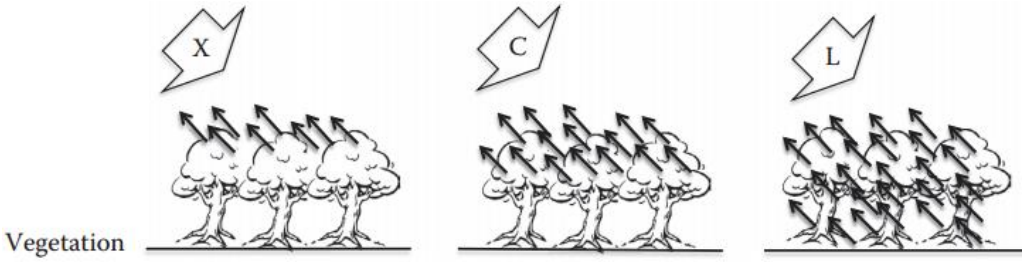



**Figure 1.** Vegetation penetration capabilities of X, C and L radar bands (data source: Pohl. 2017)

Polarimetric SAR sensors are capable to transmit and detect the vertical (V) and horizontal (H) components of the backscattered radiation (Pohl C., 2017). Hence, there are four possible polarization configurations: co-polarization- HH, VV, cross-polarization- HV, VH, and their combinations (Aulard-Macler M., R. Barstow, 2011). Polarimetric SAR backscattered energy is directly depending on the physical properties of the vegetation elements and influences on backscattering mechanism (Pohl C., 2017). The rough surface significantly depolarizes the signal in a greater magnitude while smoother surface depolarizes the signal at the lower magnitude. Thus, degree of polarization can provide valuable structural information about the forests (Periasamy, 2018). Laurin and Santoro in their articles affirm that cross-polarization has priorities over co-polarized backscatter in forest biomass information retrieval (Laurin et al., 2018; Santoro & Cartus, 2018). A number of studies point out the VH backscatter from Sentinel 1 C-band and HV backscatter form Alos Palsar and Alos2 Palsar2 L-band SAR to have more reliable correlation with AGB data over the co-polarization and other polarization configurations (Huang et al., 2018; Laurin et al., 2016)

Literature review shows several attempts of image texture characteristics implementation for biomass estimation using optical imagery back in the past (Eckert,

2012). In the more resent literature review by Santoro & Cartus (2018) the investigation of texture characteristics and polarization decompositions of SAR imagery are bringing smaller retrieval errors in forest AGB estimation compared to only backscatter values. Papers published by Berninger, Thapa and Huang depict improved correlations between image texture (namely, homogeneity, contrast, entropy, volume scattering) and the field data from allometric calculations (Berninger et al., 2018; Huang et al., 2018; Thapa et al., 2016). In case of SAR data, the texture is a measure of the spatial homogeneity of the backscattering and describes the properties of surface, such as smoothness, regularity and tonal variation, and so, should enclose information about forest structure (Santoro & Cartus, 2018). Chen et al. also show that Sentinel 1 image textures are the most relative and important predictors for AGB estimation compared to the original backscattering data (Chen et al., 2018). This review shows that the implementation of texture methods for forest biomass retrieval are understudied yet and seem to have high potential in improving the estimation accuracy.

### 2.2.1 Combination of Multisource Data for AGB Monitoring

Combination of SAR and optical data is being intensively implemented mostly in the last couple of years and shows high potential to improve the forest biomass prediction accuracy (Joshi et al., 2016; Kumar et al, 2016; Laurin et al., 2018). Joshi et al 2016, in his review of existing literature finds that different SAR data with conjunction of optical data have comparatively higher accuracy than SAR and optical data alone (Joshi et al., 2016).

Besides the prediction accuracy improvement, combination of globally available optical sensors such as Landsat 8, Sentinel-2, and SAR sensors like Palsar-2 and Sentinel-1 with high temporal resolution, is becoming a very important tool in a way that ensures operational and continuous global forest cover monitoring in consistent and robust manner (Reiche et al., 2016). Joshi et al. (2016) and Reiche et al. (2106) assume that operating on different physical principles, hence, the SAR and optical sensors provide synergistic information on the forest properties. As we already saw, the first is dependent on the size, orientation and dielectric properties, density, while the second is dependent on the leaf structure, moisture and pigmentation. This lets them consider that combination of those different satellite imagery should be

promising approach for increasing the accuracy of AGB estimation in forests (Joshi et al., 2016; Reiche et al., 2016). There is growing focus on this approach and the majority of the researches reviewed are from 2018 (Chen et al., 2018; Kumar et al., 2016; Laurin et al., 2018; Periasamy, 2018; Vafaei et al., 2018). The most often used optical sensor is Sentinel-2 because of the high resolution, free availability and the global cover with high frequency revisit cycle. Laurin et al. in their paper (2018) showed that combination of Sentinel-1 and Sentinel-2 strongly improved the forest AGB estimation in Mediterranean shrublands (around 14% with respect to the sensors separately). As an input to the prediction model all the bands form Sentinel 2 as different variables were used. Chen et al. (2018) generated number of vegetation indices and biophysical variables as input variables to the prediction models. The results show that the vegetation biophysical variables are outstanding compared to the other Sentinel-2 products while combining with Sentinel-1 products (Chen et al., 2018).

Laurin et al (2018) and Chen et al (2018) highlight that combination of multi-source satellite imagery improves the saturation level for forest AGB estimation. Laurin et al in their research (2018) got up to 400 tones/ha accurate estimation combining Sentinel 1,2 and Palsar 2 satellite imagery. Zhang assumes that interferometric phase and coherence methods have potential to overcome the saturation in some extends (Zhang et al., 2018).

### 2.2.2 SAR Drawbacks

There are number of challenges while analyzing and interpreting SAR images for land applications and particularly for forestry. Three main drawbacks of SAR data can be separated that severely affect the measurement accuracy, namely, speckle noise, the bias on the backscatter value because of the mountain relief and saturation of electromagnetic signal.

**Speckle noises**: The major problems are the speckle noises on SAR images that might cause for poor classification (Joshi et al., 2016). Unlike the optical remote sensing, radar scanning is coherent interaction of the signal with the surface objects. As the result of coherent summation of the signal scattered form, the ground scatterers has random distribution within each pixel and is called speckle noise (CRISP 2001, n.d.).

To reduce speckle noise in the image multi-looking processing property is implemented. This method helps in smoothening and noise reduction on the image, nevertheless it compromises the resolution of output. The more looks, the lower the resolution (Aulard-Macler et al., 2011; Bruniquel & Lopes, 2010). The number of looks is rarely more than 3 or 4 for satellite data (Grandi, et al., 2004). Several approaches such as multi-look summation, single-look complex summation, multi-look intensity summation and other methods developed by Lee, Boxcar, Bruniquel and others are available in toolbox ("Broxcar", "Lee", "Refined Lee", "Lee sigma" etc.) in the free software of SNAP 6.0.5 developed by European Space Agency (ESA).

**Mountainous terrain:** Another limitation of SAR is the sensitivity of backscatter to the mountainous terrain. Due to the side-looking mode, the backscattering from elevated objects leads to foreshortening, layover and shadows on the image intensity when they are projected to ground-range images (Pohl C., 2017). Geometry effects on the image radiometry, causes extensive distortions on the SAR imagery and, thus, reduces the forest biomass estimation accuracy on mountainous areas. DEM data is needed to reconstruct image geometry and perform geometric correction (Tian et al., 2014). The overwhelming majority of studies complete geometric correction using Shuttle Radar Topography Mission (SRTM) 30 m resolution DEM by NASA. Berninger in his study excludes slopes higher than 10 degrees to avoid from possible uncertainties on the SAR image (Berninger et al., 2018).

**Saturation:** A general limitation of forest biomass estimation by both optical and SAR imagery is the saturation of the signal depending on different condition. Those conditions are referring both the system for scanning and target properties of the object being scanned. The reason is forest density, therefore, the amount of biomass, used scanning wavelength and the moisture in the soil. Optical scanning, having weak canopy penetration capacity, contains information of forest mainly on horizontal direction, thus gets easily saturated in dense forests with close canopy (Vafaei et al., 2018; Zhang et al., 2018). SAR method has proven to be less sensitive to the forest density and get saturated at a higher level of biomass (Sinha et al., 2015). Depending on the used signal wavelength, different SAR imagery has shown different saturation level. The longer the wavelength, the stronger penetration power, thus, less saturation problems (Chen et al., 2018). Huang et al (2018), Vafaei et al (2018), Sinha et al

(2015) and other authors prove that C-band SAR backscatter gets saturated faster than L-band.

For the C-band Pariasamy (2018) notes 60-70 t $ha^{-1}$ saturation level. Berninger et al 2018 report result of 150-200 tones/ha saturation level for L-band backscatter. In addition to this, he claims that by using image texture characteristics it is possible to increase the result until 250 tones/ha. And Sinha in his review found saturation at 100-150 tones/ha for L-band. Huang et al (2018) state that saturation level also depends on forest state, type, as well as ground surface type and weather conditions.

### 2.2.3 SAR Dense Time-Series Analysis

As the literature review shown, the L-band SAR has stronger abilities for forest AGB compared to C-band, however, not many sources are freely available for L-band SAR data. The launch of Alos1 Palsar 1 by Japanese JAXA project is providing L-band images since 2007 and has huge investment in forest AGB monitoring. Alos2 Palsar2 launched in 2014 continues the same thematic focus and gives higher resolution images. Still, the temporal frequency of available L-band images is not sufficient for forest phenology monitoring (Palsar1 products are available one image per year, and Palsar2 images are just several scenes per year, even though, the last ones are not freely available) and thus, available literature for L-band time series studies for forest AGB have focus on forest change detection in long time period (Reiche et al., 2017). Berninger et al. (2018) uses 3 scenes from 2007 to 2016, Bouvet et al. (2018) use 4 scenes only from 2007 to 2010 from Palsar1 sensor). The launch of the open access NASA ISRO SAR (NISAR) Mission planned in a few years will continue growing of L-band SAR data availability (Huang et al., 2018).

The launch of C-band Sentinel-1a in April of 2014 by ESA in the frames of Copernicus mission in contribution to Free Big Data movement in remote sensing data has extremely expend the SAR data availability and potential for forest monitoring and particularly for AGB analysis (Torres et al., 2017). After launching of Sentinel-1b in April of 2016, which works together with Sentinel-1a it the same orbit, it became possible to gain C-band SAR images with 6-day repeat cycle on equatorial zone (revisit frequency is higher in higher latitudes) (ESAa, 2019). Thus, Sentinel-1 has dramatically improved spatial resolution, revisit time and coverage around the

globe in respect to its predecessors Envisat SAR and ERS 1,2 SAR amongst C-band SARs (Potin et al., 2016). The Level-1 products are getting available in 24 hours after scanning. Thereby, Sentinel-1 became a powerful tool for near real-time environmental analysis and change detections.

Despite the short time in the orbit, Sentinel-1 dense time-series have been already extensively applied in many fields and have proven to be significantly useful in agriculture and crop monitoring (Ghazaryan et al., 2018; Whelen & Siqueira, 2018), land cover classification (Balzter et al., 2015; Skakun et al., 2015), soil moisture estimation (Liu et al., 2017) and in many other thematic areas. For forestry there are several applications developed over the last decades through the information based on SAR time series. The main applications refer to ABG estimations, clear-cut detection, forest phenology and forest condition analysis. Dostalova and Rüetschi base on the statement that SAR backscatter is sensitive to forest seasonal cycles such as leaf-on or leaf-off conditions, successfully could implement Sentinel-1 C-band multitemporal time-series for forest type (coniferous-deciduous) deliniation for central end north Europe and in Northern Dakota in the US (Dostálová et al., 2018; Rüetschi et al., 2018). Based on the backscatter variation depending on the forest conditions, forest type and moisture, Laurin have used Sentinel-1 dense time series to find the best season and the best scenes of C-band images having stronger backscatter correlation with forest AGB (Laurin et al., 2018). He found backscatter form coniferous forests to be more stable during the year and higher correlated to the AGB value. Nevertheless, Huang et al. (2018) clarify that the forest AGB values derived only from C-band SAR imagery are not robust for forest operational monitoring.

Multitemporal SAR data is specifically useful for coherent speckle noise filtering. For this purpose, the Multi-temporal filtering methods are investigated. Besides the speckle filtering, Laurin et al. (2018) state that SAR time series analysis for the whole year can mitigate the saturation effect in the forest.

## 2.3 Forest AGB prediction models: Model diagnostics

The actual path of AGB analysis consist of many decision points, and the way of selection can generate uncertainty in the output result. This, one of the most important decisions for the AGB estimation analysis is the proper prediction model selection

(Dungan, 2002). Chen et al (2018) define the forest biomass prediction as a mapping process for an estimation of biomass values at the location without observation based on the points values at nearby observations and/or considering other factors at the site using various methods. Those methods he divides into two main categories: parametric (various statistical regression methods) and non-parametric (machine learning algorithms). In case of parametric models, it is easy to interpret and calculate the expression relating the dependent and independent variables, even though there is no simple global linear relationship between remote sensing data and forest AGB as far as there are many factors affecting it (Chen et al., 2018). It was also reproted that the majority of researches still use parametric models for biomass prediction. Two of the most often used methods are stepwise regression and multivariate linear regression. To handle high biomass variability and saturations of remote sensing imagery, Berninger at el. (2018) assumes multivariate linear regression to be superior to other models. Santoro & Cartus (2018) find, that if the aim of estimation is to deliver a biomass estimation that will be used as a layer for another purposes, then non-parametric models are quite performing. Instead, if the focus is to develop an algorithm that is robust to changeable environmental conditions, then the privileges should be given to linear models.

Non-parametric models include machine learning methods (e.g. k-nearest neighbor (KNN), artificial neural network (ANN), Support vector regression (SVR). Random forest (RF)) has shown better capability to identify the complex relationships between forest AGB and remotely sensed data (Chen et al., 2018; Englhart, Keuck, & Siegert, 2012). Because the field NFIs are not easy to organize and big datasets usually are not widely available, non-parametric methods show better prediction (Chen et al., 2018; Vafaei et al., 2018). The same studies prove that amongst the ML methods SVR shows the best results, even though, the difference between various ML prediction methods were not insignificant. Nevertheless, it has been assumed that non-parametric models, even if they reach to a better result, not always lead to better predictions for new datasets. This is because complex models can increase the risk for overfitting, i.e. explaining small datasets through large number of variables which can be partially redundant and lead to unreliable prediction (Fassnacht et al., 2014).

To detect the uncertainty level and evaluate the model performance by most of the authors were purposed the coefficient of determination ($r^2$) and root-mean-squared

error (RMSE). Less often mean absolute error (MAE) (Vafaei et al., 2018) was implemented together with the abovementioned metrics. Bouvet et al (2018) used Monte-Carlo cross-validation to validate the model's predictive accuracy. To avoid from multi co-linearity among the input variables Chen et al (2018) and Berninger et al (2018) use variance inflation factors (VIFs) in their regression analysis to exclude variables that are highly correlated. For regression model diagnostics Laurin has used leave-one-out (LOO) cross-validation approach (Laurin et al., 2018).

The result of the literature review declares the parametric regression prediction models to be the preferable and commonly used ones compared with the non-parametric ones. Despite the diversity of forest AGB models, not many studies focus on comparison of parametric and non-parametric models. A systematic comparison is quite urgent but also rare in the literature (Chen et al., 2018).

Following the extensive literature review we select Multilinear Stepwise Regression parametric model for our study to evaluate the multisource input variables and to predict forest AGB through that data.

# DATA AND STUDY AREA

## 3.1 Introduction to the Study Area

The study area is the NE forests of Armenia, more specifically, Noyemberyan forest enterprise, where the UNDP-GEF program is piloting by implementing NFI methods for forest carbon stock calculations for the first time in Armenia (Haywood et al., 2018). The first steps of field data collection are partially covered and the raw data is available, which was the main stimuli for study area choice.

Armenia has a territory of 29,800 km² and is described with complex topography and great range of altitudinal variation (lowest point has 375 m elevation from sea level, and the highest pick is 4095 m, average altitude is 1850 m) (Haywood et al., 2018). The country is located in dry subtropical climatic zone, but because of mountainous relief, the landscapes are vertical distribution and roughly divided into 4 main types: deserts and semi- deserts, mountain steppes, forests, thin forests and shrubs, and sub alpine meadows (UNDP-GEF, 2015).

Being located in between a junction of 4 main floristic regions (Old-Mediterranean, Near-east Asian, Iran-Turanian and Caucasian) (Moreno-Sanchez et al., 2007) Armenia is a home for a variety of (110 tree and 152 shrub) species. Currently, being included in Caucasian and Iran-Turanian biodiversity hotspots, the country presents a habitat for many endemic vegetation species. Based on the density of high vascular plants, the country is ranked among the first-place countries in the world with about 107 species per 1000 km² (Moreno-Sanchez et al., 2007; UNDP-GEF 00091048, 2015).

In Armenia, the forest cover extend is estimated to be over 332,000 ha (Rio+20, 2012), and ~62% out of that is located in the NE Armenia. The forest type in the north east are subtropical (61%) and temperate (39%) (MONGABAY, 2019) and are described as broadleaf deciduous mixed forests. Dominant tree species are oak (Quercus spp.), beech (Fagus orientalis), hornbeam (Carpinus betulus) and oriental hornbeam (Carpinus orientalis) (Moreno-Sanchez et al., 2007). The NE forests are located in two provinces (Tavush and Lori) and consist of 10 forest enterprises and “Dilijan” National Park occupying over 253,000 ha area, in which 215,000 ha is forest

covered (UNDP-GEF, 2015). Annual precipitation is 500-540 mm and the climate is described with warm and dry summers and temperate winters (FMP, 2018).

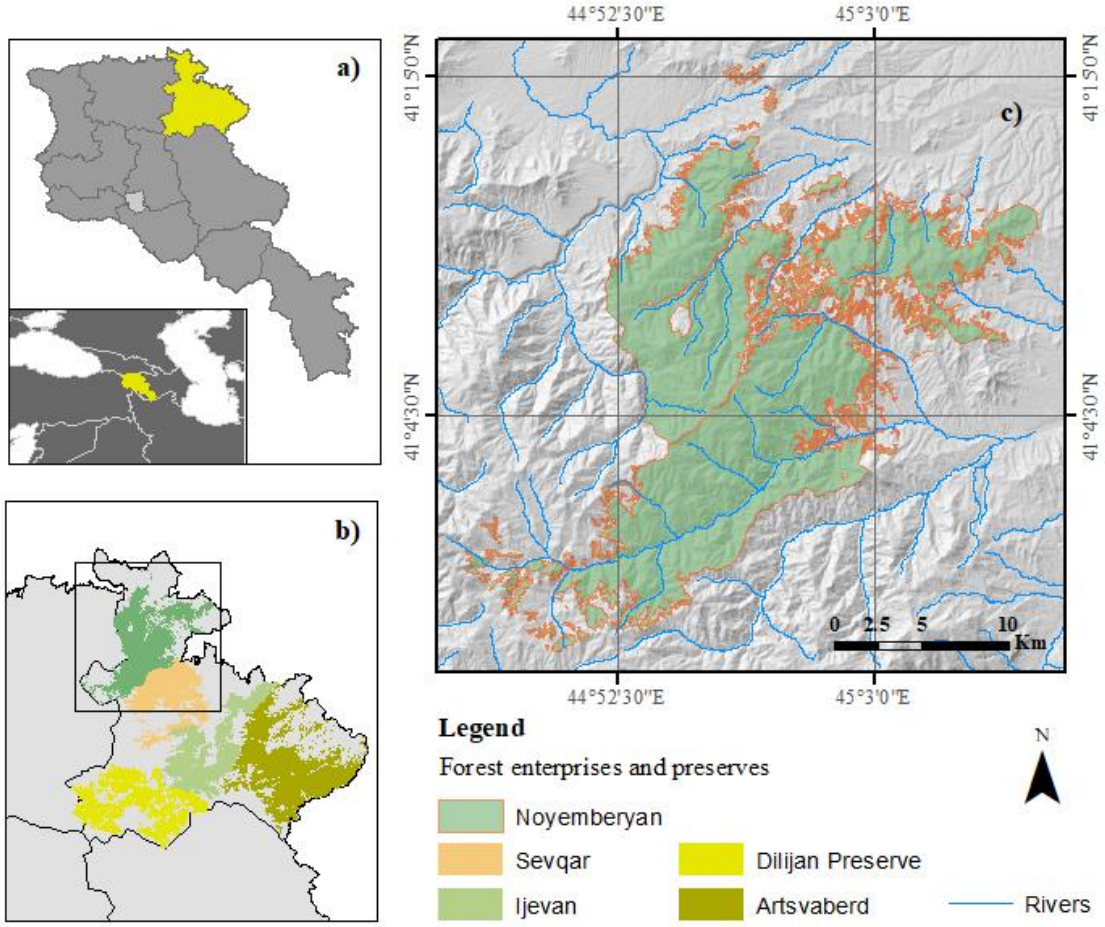



**Figure 2**. Location of the study area. a) Location of the Tavush province, b) forest enterprises in Tavush province, c) "Noyemberyan" forest enterprise

This research focuses on the forest sites located in Tavush province (Artsvaberd, Ijevan, Sevqar, Noyemberyan and "Dilijan" National Park), The main focus is Noyemberyan forest enterprise for several reasons:

- The pilot project for national forest carbon inventory has been implemented in Noyemberyan forest enterprise and can be as a ground base for evaluation of results,
- The field sampling density is much higher in Noyemberyan forest enterprise (55 sampling points out of 115),
- Some of remote sensing data has huge distortions over the other forest sites which practically makes those data not useful for forest monitoring on the whole area,
- All the forests in the NE Armenia have very similar characteristics in terms of forest type, structure, amount of biomass, terrain conditions, and so, we do not put

uncertainty while using the field sampling data from one site for model training and applying on the other one.

For statistical analysis and model training information was collected from all the 5 forest sights but the project's main focus area is considered the forests of Noyemberyan. Therefore, all the objectives and hypothesis are applied and the final maps and evaluation are done for Noyemberyan forests only.

The forests in Noyemberyan enterprise is 29334 ha and is located in between 600-1850 MASL altitudes and the distribution is as following: <800m – 4.2%, 801-1200-47.9%, 1201-1600-37.8%, 1601> - 10.1%. Forest structure, quality and dominant species vary based on the altitude - spread low density in the low altitudes to high density in the middle. The topography is highly fragmented and remarkable with steep slopes and aspect composition variations. This strongly affects the forest distribution. There are more forests located on the north slopes (58.6%) than on the south slopes (41.4%) (FMP, 2018).

## 3.2 Introduction to Data

### 3.2.1 Field Data

Field data are gained from the UNDP-GEF ongoing project. The initial number of plots were 115, out of which 55 were from Noyemberyan enterprise collected in 2017, and the rest are from the other forest sites of Tavush province collected in the beginning of 2018. The field sampling plots' locations are chosen based on a systematic sampling plot design (**Figure 3**).

| **Strata** | **No of samples** |
|---|---|
| Beech | 18 |
| Hornbeam | 22 |
| Oak | 18 |
| Pine | 12 |
| Other | 19 |
| Disturbed | 26 |
| **Total** | **115** |

**Table 3.** Field sampling data per strata

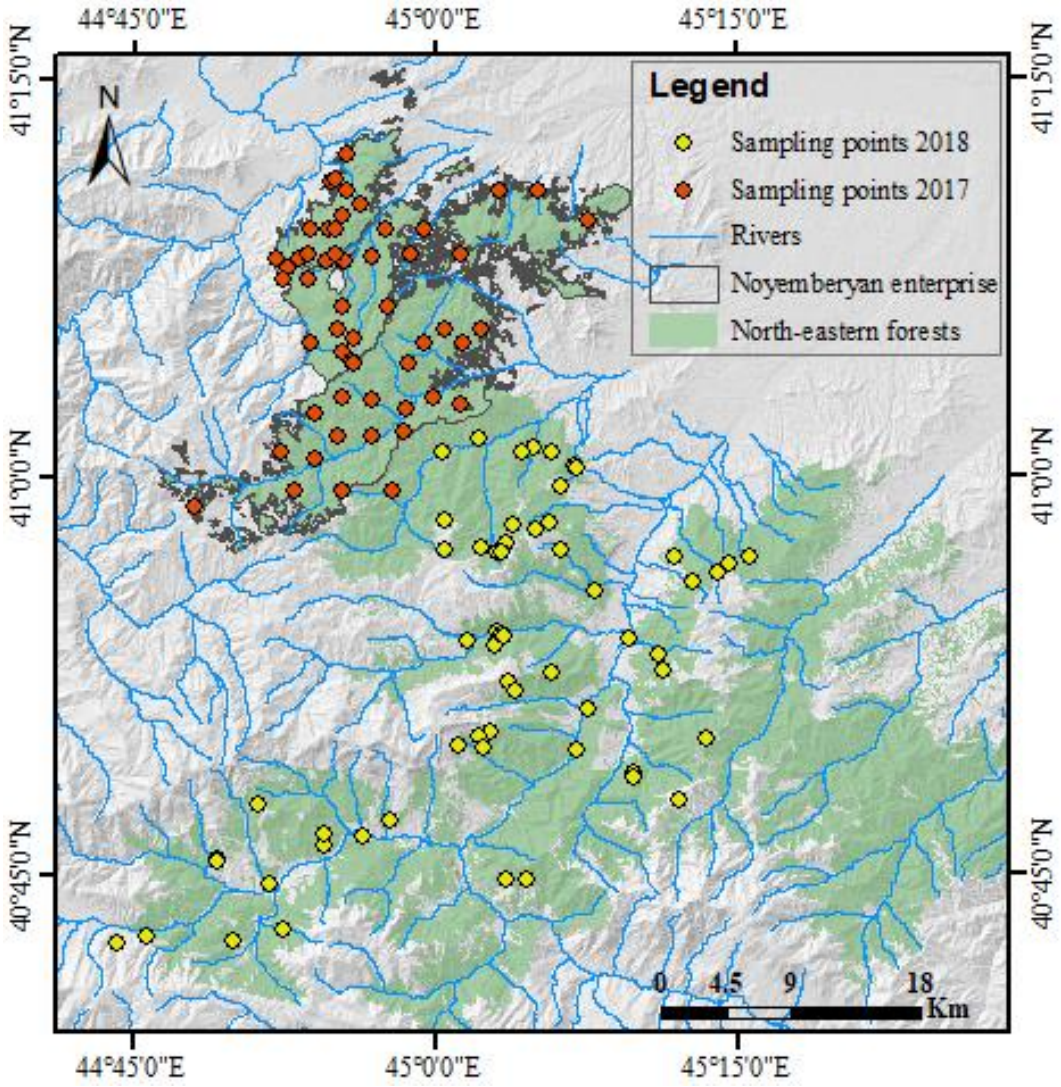



**Figure 3.** Field sampling plots distribution

There was also used a forest stratification map to ensure that the sampling plots include all the forest types (**Table 3**). Fixed size of 0,1 ha circular plots were established with 17.84 m radius (depending on the slope degree it might change in order to ensure a 0,1 ha plot size when projected on a plane). All the large trees (>8 cm in diameter) are assessed for the following properties: species, DBHOB, tree status, decay class, crown class and many other parameters, that are not relevant to this study, thus will not be reported.

### 3.2.2 Sentinel 1 Data

For this study Sentinel 1A and Sentinel 1B C-band SAR data from Copernicus project of ESA was used. The data is interferometric wide swath (IW) scanning mode, with

250 km swath width. We used Level-1 Ground Range Detected (GRD) product which is already Multi-looked (one look in range and five in azimuth), Geocoded and it is available with 10x10m pixel spacing both in Copernicus Open Access Hub[1], (the online system of the ESA), and in Google Earth Engine JavaScript API[2] (GEE). Both satellites orbit in near polar, sun-synchronized orbit at 693 km altitude and in the same orbital plane (Torres et al., 2017). Sentinel 1 C-band SAR has 5.405 GHz frequency (corresponding to a wavelength of ~5.6 cm) and provides images in two polarization modes: VV co-polarization and VH cross-polarization.

The incidence angle is between ~31 and ~46 degrees. The resolution was set to 10 m. Each image contains 3 bands: two for backscattering intensity (VV, VH) and one for incidence angle. Because of heterogeneity in the parameters of available images there is a need to filter to down the data for a homogeneous subset. In order to have the same incidence angle for all the images the same relative orbit number 152 was selected. There were 72 scenes acquired for the whole year of 2017 (30 from Sentinel 1A, 42 from Sentinel 1B), 6 images per month in average.

For Sentinel 1 data available in GEE platform the backscattering coefficient from natural values of sigma naught ($\sigma^\circ$) is converted into dB values.

$$\sigma^\circ = 10 * \log_{10} DN \quad (3)$$

Where the DN is the digital number of the natural values.

($\sigma 0$ (dB) = $10 * \log_{10}(\text{absolute } (\sigma 0))$).

### 3.2.3 Sentinel 2 Data

The Sentinel 2 imagery was obtained from 2 satellites: Sentinel 2A and Sentinel 2B. Data is available free of charge in Copernicus Open Access Hub and GEE platform. Sentinel 2 data are characterized by13 spectral bands with 10-m, 20-m, and 60-m spatial resolution and a radiometric resolution of 12 bit. Two form of data are available: Level 1C (L1C)– top-of-atmospheric reflectance product, and Level 2A (L2A) – bottom-of-atmospheric reflectance product.

All the acquired images are from 2017: June and July (16 L1C products with cloud cover less than 15%), September (8 L1C products with cloud cover less than 25%) and October (2 L1C products with cloud cover less than 10%). Only 10m and 20m

resolution bands were used of the analysis: B2 (490nm), B3 (560nm), B4 (665nm), and B8 (842nm) 10m spatial resolution bands, B5 (705 nm), B11 (1610nm), and B12 (2190nm) 20 m spatial resolution bands (ESA, 2018).

### 3.2.4 Alos Palsar Data

Global Palsar-2/ Palsar L-band SAR data were accessed at a 25 m scale in GEE platform. The dataset is generated by applying Japanese Aerospace Exploration Agency's (JAXA) processing and analysis technique to a lot of images obtained with Japanese (Palsar and Palsar-2) radars on Advanced Land Observing Satellite (ALOS and ALOS-2) satellites/carriers (JAXA, 2018).

Palsar/ Palsar-2 images are L-band SAR (~23.5 cm wavelength) and the images are acquired in Fine Beam Dual polarization (FBD, 70 km swath width) mode: HH- horizontal transmit, horizontal receive, HV- horizontal transmit, vertical receive. The incidence angle is between 28.6 degree and 32.9 degree (CEOS, 2016; JAXA, 2018). The temporal interval of the images contained by the mosaic is generally 1 year, and no information is available about the scanning date for a particular area. According to the tutorial of Global 25 m Palsar product, the images are selected taking into consideration the weather information, in order to avoid radar backscattering saturation effect caused by the moisture (JAXA, 2018).

Ortho-rectification and topographic corrections on the SAR data are applied using 90m SRTM DEM, which is not preferable for this research, as out study area is highly fragmented and this can cause for important information lose.

Backscattering data are stored in digital number (DN) of unsigned 16 bit. These values can be converted to gamma naught ($\gamma^\circ$) natural values in decibel unit (dB) by the following equation:

$$\gamma^\circ = 10 * \log_{10}(DN^2) - CF \quad \text{(dB)} \qquad (4)$$

Where CF is the calibration factor and for product of Palsar/ Palsar-2 is measured to be ~83.0 dB.

### 3.2.5 Forestry Ancillary Data

The forestry ancillary data used in the scopes of this research are the forest type information from forest stratification map, and the topographic parameters of the sampling plots, namely, the aspect and the slope information. Though this information is available from the forest inventory data, we used the necessary equivalent information retrieved from DEM. The SRTM DEM which was gained from U.S. Geological Survey portal (USGS, September, 2018) free of charge. DEM data is low spatial resolution (30 m).

# APPROACHES, METHODOLOGY, DATA PREPARATION

## 4.1 Approach and General Methodology

Different pixel-based methods were used in order to answer the research questions and achieve the objectives. The pixel size is chosen 30m quadrat to be equal to the SRTM DEM product, which was used for terrain corrections. Also, certain circumstances were considered: the minimum mapping unit (MMU) is 0,09ha, which is competitive with the area of field sampling plots (0,1 ha), and after applying a scaling factor of 0,9 on the field ABG estimation, we replaced the circular sampling plots with squared pixels. Projection system for the whole project was selected WGS-84 38N local for the study area projection. All the rasters were aligned in order to match the pixels for raster analysis.

Field data was carefully examined as certain conditions should have been ensured before using SAR data for regression analysis. Even if a part of the field sampling data were from the beginning of 2018, we kept using the remotely sensed data from 2017. As the annual growth of wood for the forests in Armenia is estimated as 1.4 $m^3$/ha (Rio+20, 2012), which we consider to be not significant for this study.

Literature review shows, that working with SAR data demands deeper understanding of the physics behind it and more careful pre-processing and processing of the data, which is crucial for improving the quality and retrieving the relevant information from SAR backscattering. Implementation of some typical processing steps, such as radiometric correction and speckle filtering are required even if we are working with images form the same sensor but with different time stamps (Navarro et al., 2016). Even more, within the scope of this study we combined data from different sensors, which assumes more pre-processing steps such as identical backscattering naught retrieval.

As the study area is suffering from complex and highly segmented topography and taking into consideration the fact that in such case even the radiometric terrain correction is not sufficient for mitigating the bias and the gain on the SAR backscattering data caused by high topographic variation (Pohl C., 2017), we implemented ancillary topographic information of the sampling sites directly into

regression model. Simple removal of the steep slopes and masking of the foreshortening and shadows on the SAR data is not possible, as those areas are occupying the most part of the study area.

Besides the tree covered areas, forest sites include also non-forested forest lands (meadows, pastures, herbs and forest disturbance areas) are sometimes significantly big and can mislead the forest biomass estimation. For the FNF map generation in purpose of precise delineation of forests, Sentinel 2 imagery and Random Forest (RF) classification method was used. This non-parametric classifier was selected since it does not require any statistical analysis of input variable and can handle with multicollinearity effect.

Recent Studies show successful implementation of Haralic texture analysis (GLCM matrices) for biomass prediction, when different parameters of the local variance of the pixel values is calculated (Chen et al., 2018; Huang et al.,, 2018). Image texture analysis technique was also applied by choosing the more relevant texture parameters.

For the biomass prediction and AGB mapping backward Stepwise Multiple Linear Regression (SWR) model was used to automate the best explanatory variable selection. Even if the studied state some better results for non-parametric models over the parametric ones for biomass prediction, the use of SWR allows us to evaluate and compare each of the input variables (i.e. the input data and the environmental conditions), which is one of the main focus areas for this research. Variables with parameters of p-value $< 0.05$ and VIF $> 5$ were excluded form regression model (Berninger et al., 2018; Laurin et al., 2018).

The **Figure 4** provides the flowchart of the overall methodology (The complete flow chart for the methodology see in **APPENDIX B**). The flow chard is designed following way: raw input data (gray), information of unprocessed data (blue), data pre-processing and processing steps (white), processed-ready for analysis data (green), classification and regression analysis (brown). Red dashed boxes are indicating the main sub-processes which will be discussed in a detailed manner in the next chapters. Those sub-processes are:

- Field data calculation
- Sentinel 1 data pre-processing
- PALSAR data pre-processing

- GLCM texture analysis
- Sentinel 2 data pre-processing
- Ancillary data pre-processing
- Forest/Non-Forest classification

The regression analysis, model diagnostics and evaluation are discussed in the **Chapter 5.**

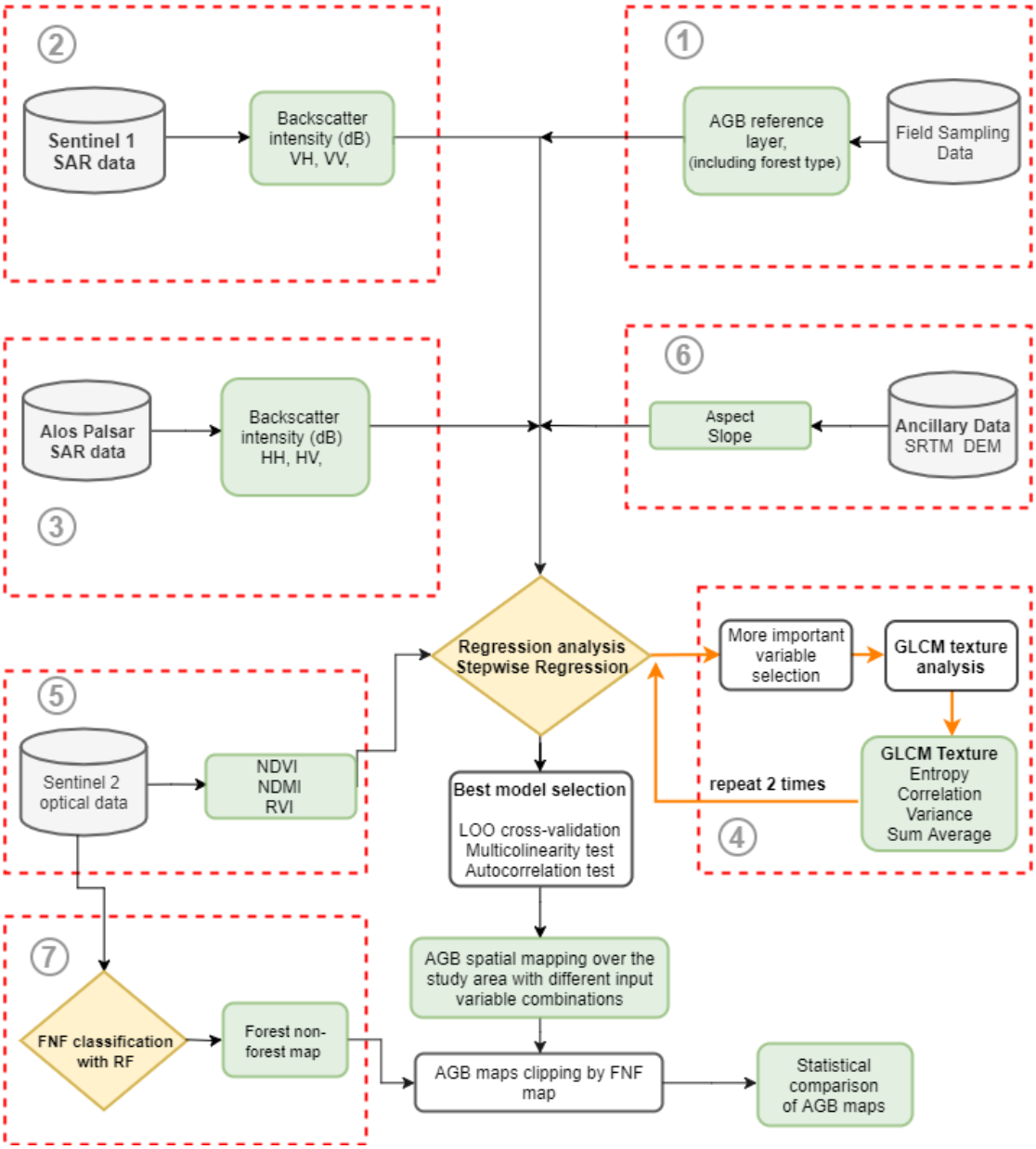



**Figure 4.** Flow chart of general methodology

## 4.2 Tools

The tools used for the data pre-processing and processing are presented in the **Table 3**. The focus is on the open software and free of charge cloud environment. Satellite image pre-processing and processing is mostly carried out in the Google Earth Engine cloud-based environment by using JavaScript API.

| Process | Tools |
|---|---|
| S1 pre-processing<br>Palsar mosaic pre-processing<br>S2 vegetation indices calculations<br>Hansen Global forest map implementation<br>FNF classification with RF<br>Classification accuracy assessment<br>GLCM texture analysis | GEE JavaScript API |
| S2 L1C to L2A processing<br>S2 vegetation indices calculations | SNAP, Sentinel Toolbox, Sen2Cor plugin |
| Regression analysis<br>Cross-validation and model diagnostics<br>AGB mapping on the study area | R |
| Projection transformation<br>Raster re-scaling and aligning<br>Field sampling data processing<br>Shape to raster conversions | QGIS |
| Map designing<br>Aspect and Slope processing | ArcGIS Desktop |
| Field data calculations | MS Office Excel |

**Table 4.** Tools and related processing

GEE is cloud based geospatial processing platform and provides huge computational power, which made possible the processing of enormous amount of satellite imagery and conduct dense time-series analysis. GEE is an environment for planetary-scale environmental data analysis and contains the archives of many publicly available remote sensing imagery (Gorelick et al., 2017).

## 4.3 Field Data Calculation

As it was mentioned in **Chapter 3.2.1** forest inventory raw data was initially available from 115 sampling plots done by forest systematic sample plot design. The field data calculation steps are presented in **Figure 5**:

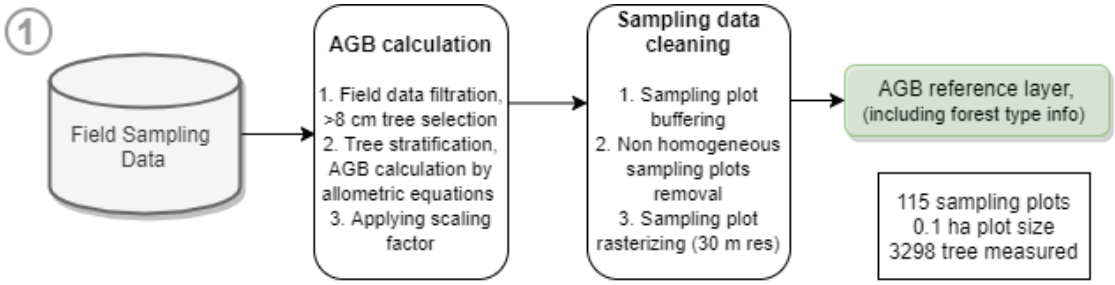



**Figure 5.** Flow chart of field data pre-processing steps

Amongst different vegetation parameters measured during the inventory, for aboveground biomass calculation we used DBHOB parameter of the trees and simplified allometric equations offered by manual for the GEF's pilot project. The species specific allometric models are available for each of the main tree species and one general equation for the rest of species. Those allometric equations are developed for the tree species common for the region (Shahrokhzadeh et al., 2015) and are presented in **Table 5**.

| Tree species | Allometric equations | No of eq. |
|---|---|---|
| Beech | $AGB_{Live} = 0.353 * D^{2.191}$ | (5) |
| Oak | $AGB_{Live} = 0.0743 * D^{2.012}$ | (6) |
| Hornbeam | $AGB_{Live} = 4.046 * D^{1.1599}$ | (7) |
| Pine | $AGB_{Live} = 0.353 * D^{2.191}$ | (5) |
| Other | $AGB_{Live} = 0.353 * D^{2.191}$ | (5) |

**Table 5.** AGB allometric equations for dominant tree species

In the equations D stands for diameter at breast height over back (DBHOB) of the large trees in cm. Those equations are established for aboveground living biomass calculation. As the data explorations showed, the amount of standing dead trees are very few per plot, also, from the perspective of SAR scanning, that circumstance doesn't practically affect on the SAR backscattering generation, thus, we used those formulas also for calculation of dead standing tree biomass.

Once the AGB is calculated for each plot, a scaling factor of 10 was applied to calculate the amount of biomass per ha (AGB_Live*10). 3298 large trees were measured and 1701 t AGB was calculated from all the sampling plots. The maximum of 680 t $ha^{-1}$ biomass was observed and the average standing biomass in the study area is 148 t $ha^{-1}$.

SAR data records realistic backscattering data over the homogeneous forests. Not homogeneous distribution of forests can mislead in backscattering values (Huang et al., 2018; Laurin et al., 2018). Hence, two steps data cleaning was done to detect the sampling points with not homogenously forested areas. First, a minimum threshold of 2 t/plot on the biomass was established assuming, that less than that amount will not have homogeneous tree cover. This way 22 samples were removed from the beginning (17 out of this 22 had no large tree biomass at all).

For the second step we use the technique of buffering implemented by Huanag et al. (2018) and Laurin et al. (2018). The logic behind making buffers around the sampling plots is that after converting the point shapefiles (centers of sampling plots) into raster, in most cases the points do not match with the center of squared pixels of raster. This is due to the systematic sampling grid size, which was different from the grid size of our raster (30x30). Therefore, the pixels are randomly allocated towards the centers of the sampling plots. Couple of tries to align the raster with the points gave different dispositions of pixels. Thus, we assume that a pixel representing the sampling point can gain the satellite image value from the surrounding of the sampling plots by unpredictable way (**Figure 6).** To ensure that the backscattering intensity represents the real field sampling AGB value and does not mislead the backscattering statistics because of non-homogeneity in the forest cover, we created buffer of 42m (the maximum distance the pixel can be situated). The purpose was to audit the similar forest structure and similar reflectance for the adjacent sites of each plot. Visual checkup of forest homogeneity within the buffer areas was complete using ESRI basemap imagery and Google Earth high resolution imagery. All the inaccurate sampling points were excluded from further calculations (**Figure 6**). This way more than 11 samples were removed.


**Figure 6.** Field sampling points with AGB. Green polygon shows the sampling design, red polygon is the buffer zone, pixels represent field sampling plots as a raster, numbers show the amount of biomass for the sampling plot in kgs

## 4.4 Sentinel 1 Pre-processing

As it was indicated in the SAR data description in the general methodology (**Chapter 4.1**), depending on the purpose of use, SAR data should be pre-processed, in order to retrieve the needed information. In particular, besides the typical corrections suggested for SAR data, in this study gamma naught was used (gamma-zero, $\gamma^0$), which is terrain-corrected with the ellipsoid model and does not depend on the incidence angle (Small, 2011). This pre-processing step is of high importance for this study, as the study area is notable for its complex topography. The detailed steps of pre-processing are presented in **Figure 7**.

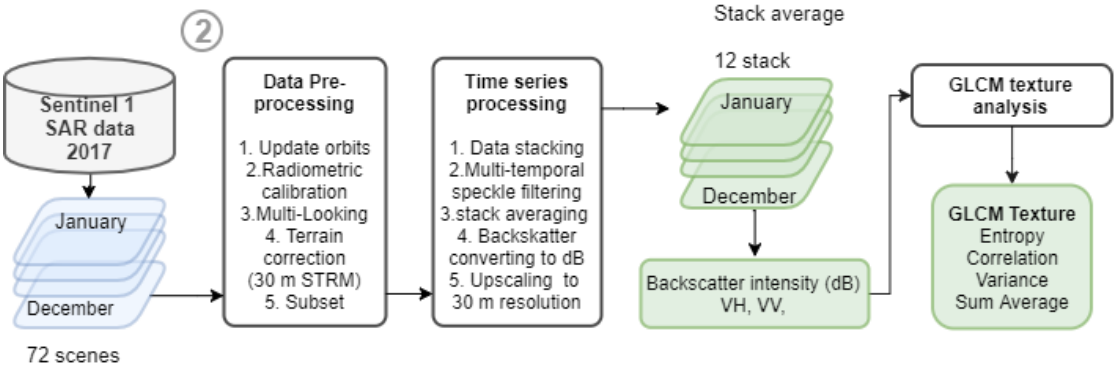



**Figure 7.** Flow chart of Sentinel 1 pre-processing steps

Sentinel C-band data analysis were carried out in GEE cloud environment, where the GRD data is available with some pre-processing steps being applied. Those corrections are: orbit update, radiometrically corrected to sigma naught ($\sigma^0$), natural value conversion to dB using equation (1), geometric terrain correction (Range Doppler Correction) offered by Sentinel Toolbox od ESA using SRTM DEM (GEE API documentation, 2019).

In order to fulfil the further pre-processing steps, we created and applied automated Sentinel 1 time-series processing chain. First of all, the values from dB were converted back to their natural values. To retrieve the gamma-naught, we used formula (6):

$$\gamma^0 = \frac{\sigma^0}{\cos\phi} \tag{8}$$

where $\phi$ is the incidence angle (Small, 2011).

12 stacks for the whole year were composed using the scenes within each month. Refined Lee speckle filter was applied on each stack for multitemporal speckle filtering in order to reduce the speckle noise (Yommy et al., 2015). For this 7x7 kernel size was set up. Afterwards, the monthly stacks were averaged with the statistic mean value and normalized to dB. In the end, all the monthly stacks were stacked into one image with 12 bands and upsampled to 30x30m pixel spacing.

All the above-mentioned steps were done for VV and VH polarization individually. For the VV polarization images -25 dB of minimum threshold was setup, and VH polarization there was -30 dB threshold setup to avoid form extreme abnormal backscattering values. This was done after careful observation of backscattering on forested and non-forested areas (Shimada et al., 2014). In total, 24 images were

generated, which were used as input variables in the regression analysis. The codes for all the above-mentioned pre-processing steps are available in GEE API and can be found in the **Appendix A**.

## 4.5 Palsar Pre-processing

Preprocessing of Alos Palsar product does not assume many steps since many of the systematic attributes are missing from the data. The conversion of DN to dB is done by applying the equation (2). The speckle filtering and extreme pixel value masking is done in the same way as was implemented foe Sentinel 1 C-band imagery (**Figure 8**). For the Palsar product we had 2 images (HH and HV) stacked into one as two different bands. Images were resampled to 30m spatial resolution.

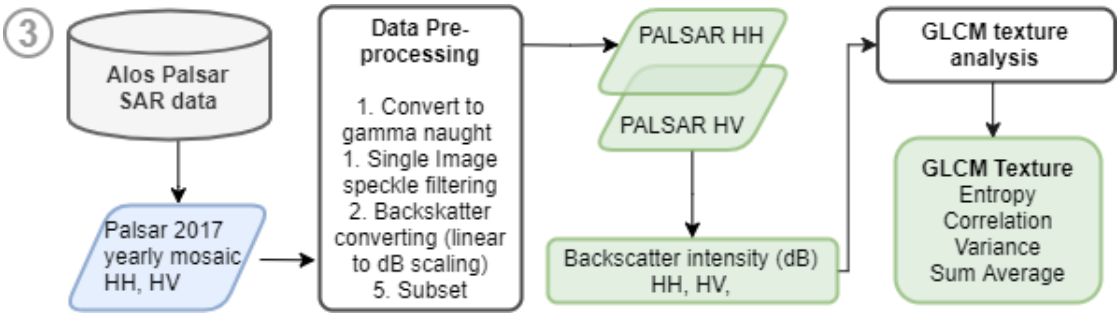



**Figure 8.** Flow chart of Sentinel 1 pre-processing steps

While pre-processing Palsar data, systematic errors were discovered on the imagery expressed in a very high pixel value. Those inaccuracies cover large and continuous spaces on the study area. Particularly, because those errors were out of Noyemberyan enterprise area, it made possible for Palsar data still to be used for the final AGB mapping.

For the Palsar mosaic for non-forest areas a noise threshold of −34 dB was set up following the tutorial for global forest map product (Shimada et al., 2014). All the pre-processing steps are carried out in GEE and can be found in the **Appendix A**.

## 4.6 Sentinel 2 Pre-processing

Sentinel 2 (S2) data were pre-processed in order to get NDVI, NDMI, RVI vegetation indices for regression analysis and for FNF classification (**Figure 9**).

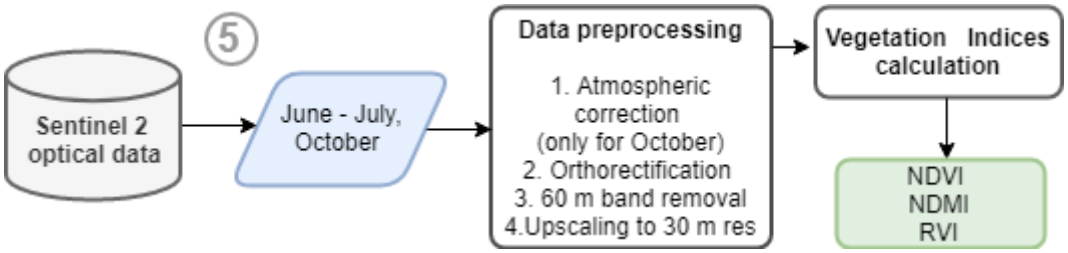



**Figure 9.** Flow chart of Sentinel 1 pre-processing steps

As was described in **Chapter 3.2.3** S2 imagery is already available as L1C product. To pre-process S2 data from August into L2A bottom-of atmospheric reflectance product Sen2Cor plugin from Sentinel Toolbox was used (implemented in SNAP software). Afterwards, the images were mosaiced and vegetation indices NDVI and RVI (equations 7 and 9) were calculated again using SNAP software.

The rest of the images were pre-processed in GEE cloud environment without applying bottom-of-atmospheric correction. The images were cloud masked, the bands B2, B3, B4, B8, B5, B11 and B12 were kept, the rest was dropped out from the future processing. All the scenes from June-July were reduced with the mean value, then NDMI index was calculated (8).

$$\mathrm{NDVI} = (\mathrm{NIR} - \mathrm{R})/(\mathrm{NIR} + \mathrm{R}) \quad (9)$$

$$\mathrm{NDMI} = (\mathrm{NIR} - \mathrm{SWIR})/(\mathrm{NIR} + \mathrm{SWIR}) \quad (10)$$

$$\mathrm{RVI} = \mathrm{NIR}/\mathrm{R} \quad (11)$$

S2 data from September was pre-processed for FNF classification. Cloud masked data was stacked and reduced with the mean value of the pixel, NDVI was calculated and added to the stack as another input variable for the classification analysis.

In the end, all the scenes were upsampled to 30m pixel size to align with the rest of the rasters for further analysis.

## 4.7 Ancillary Data Pre-processing

The aspect and slope forest ancillary data were generated from SRTM 30m spatial resolution DEM. For these 4 scenes of DEM from 2014 (updated in 2015) was used (**Figure 10**).

Because the aspect data is categorical, we combined it into 8 segments, each group consisting of 45 degrees of aspect. Completely flat areas consist 0.8% of the total are, so we disregarded that and included in the 1th segment. The numbering is designed stepwise. Afterwards, we implemented one-hot-encoding technique to transform the categorical variable into 8 different input variables with numerical values using R (R Core Team, 2018). Thus, we made it possible for this information to be used in linear regression analysis as an input variable. “Encoding” package was used for this.

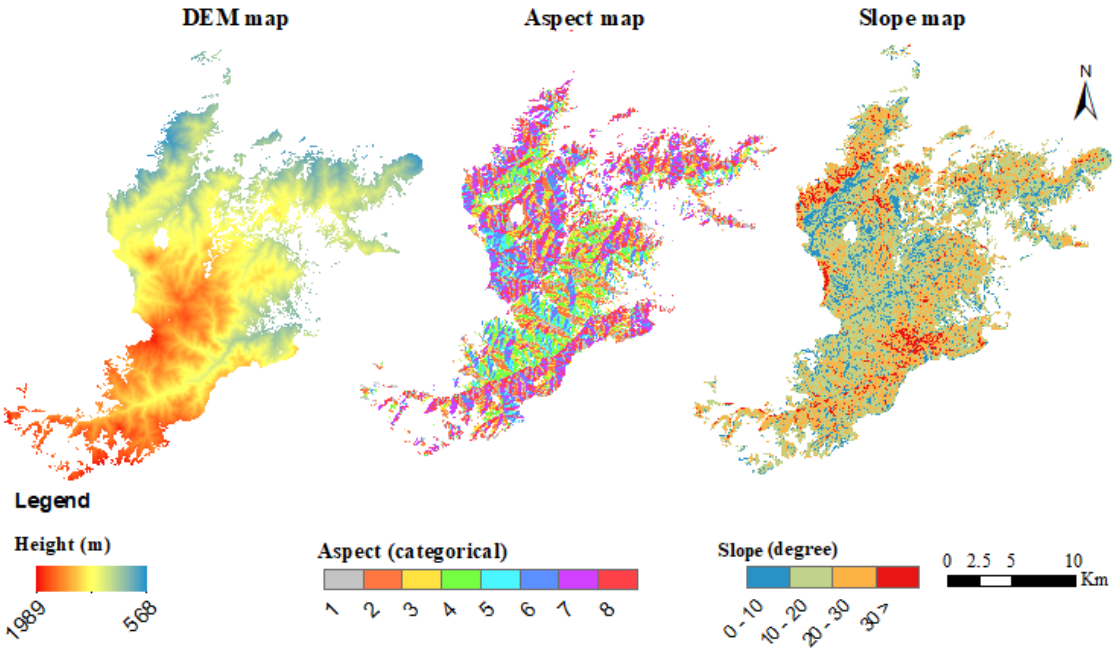



**Figure 10.** DEM, Aspect and Slope maps of the study area

## 4.8 GLCM texture analysis

Image textural features, developed by Haralic et al. (1973), measure the spatial homogeneity of the backscattering and contain information about forest structure. Several studies have proven that GLCM texture from SAR can be better describing the biomass distribution than SAR backscattering itself (**Chapter 2.2**). GLCM analysis were applied only on the SAR imagery with a selective approach (**Figure 11**), aiming to decrease the number of input variables for the regression analysis.

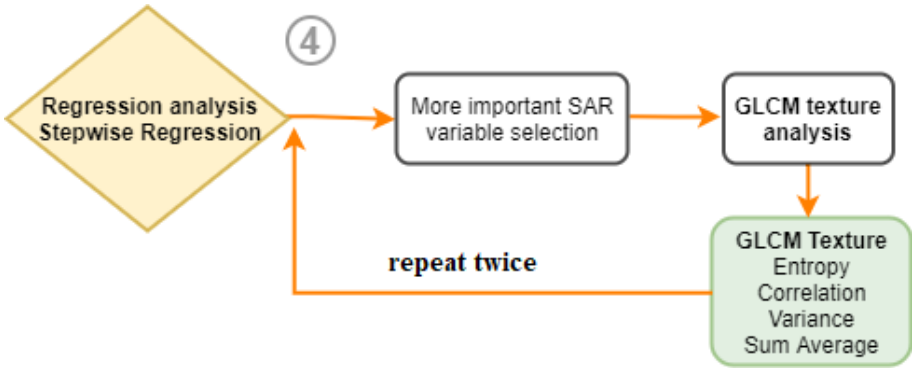



**Figure 11.** Flow chart of GLCM texture analysis steps

In order to determine the more relevant SAR variables, the regression model was run with the input variables of S1 (22 variable from time-series stacks for each VV and VH polarization), Palsar (2 variable for HH and HV polarization), Vegetation indices (3 variable – NDVI, NDMI, RVI) and ancillary data (3 variable - tree species type, aspect, slope). More important variables were selected based on their correlation significance according to the stepwise linear regression test (SWR). In the first iteration the more significant SAR variable (Palsar HV) was selected, then 14 texture measures were calculated for Palsar HV in GEE (**Angular Second Moment, Contrast, Homogeneity, Correlation, Variance, Sum average, Sum variance, Entropy, Sum entropy, Difference entropy, Energy, Difference variance, Difference entropy, Maximum correlation**). Those masseurs as input variables were added to the same variables for another iteration of regression.

This time the best texture measures with highest correlation with the biomass were determined again based on the variable significance from SWR test. Those measures are Variance, Entropy, Correlation, Sum of average (**Table 6**).

GLCM texture analysis were carried with GEE cloud environment. For these analysis SAR data in gamma naught (converted to natural values) are used. To have the real statistics of the image pixels, texture measures are calculated on the images before applying a speckle filter. 4x4 window size was setup for the calculations.

| GLCM texture | Formula | Description | No of eq. |
|---|---|---|---|
| Variance | $\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i-\mu)^2\,p(i,j)$ | Measures the dispersion (with regard to the mean) of the gray level distribution | (12) |
| Entropy | $-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}p(i,j)\log p[(i,j)$ | Measures the degree of disorder among pixels in the image; it is (approximately) inversely correlated with uniformity; images with a larger number of gray levels have larger entropy | (13) |
| Correlation | $\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{ijp(i,j)-\mu_x}{\sigma_x\sigma_y}$ | Measures the linear dependency of gray levels on those of neighboring pixels; it provides a measure similar to autocorrelation methods | (14) |
| Sum average | $\sum_{i=2}^{2N_g}ip_{x+y}-(i)$ | Measures the mean of the gray level sum distribution of the image | (15) |

**Table 6.** Formulas for GLCM texture measures. Formulas for GLCM texture measures. For all the equations, p(i, j) is the (i, j)-th entry of the normalized gray-level co-occurrence matrix, that means, p(i, j) = P(i, j) / $\sum_{ij}$ P(i, j), where P(i, j) is the (i, j)-th entry of the computed GLCM; $N_g$ is the total number of gray levels on the image; and $\mu_x$, $\mu_y$ and $\sigma_{x,}$d$\sigma_y$ stand for the Mean and Standard Deviation of the row and column sums of the GLCM, respectively (data source: Haralic et al., 1973).

# DATA PROCESSING. FOREST BIOMASS MONITORING

## 5.1 Forest/Non-Forest classification

Before starting the forest monitoring and biomass calculations, it's significant to distinguish the main concepts regarding forest definition, which will bring better insight into the field data we have, and on the extents of the span that should be explored within this research. As the forestry field information is provided by each forest enterprise (biggest forest site units), it is worth to define the forest enterprise for this case. According to the Forest Code of Armenia:

**Forest enterprise** – *a production unit with the aim of sustainable forest management.*

As the forest enterprise is forest-economic unit, it consists of forested and non-forested lands. According to the same Code:

**Forest Lands** - *forested lands and lands allocated or envisaged for flora and fauna protection, nature protection as well as non-forested lands allocated or envisaged for the running of forest economy*.

On the other hand, recent studies show that the non-homogeneous forested areas can cause for SAR backscattering value distortion on the image. To avoid from this kind of uncertainties, we offer to delineate the boundaries for forests only, inside the forest enterprises and consider those areas as the limit of the extents for this study. According to the same forest Code:

**Forest** - *interconnected and interacting integrity of biological diversity dominated by tree-bush vegetation and of components of natural environment on forest lands or other lands allocated for afforestation with the minimal area of 0,1 ha, minimal width of 10 m and with tree crowns covering at least 30% of the area, as well as non-forested areas of previously forested forest lands (Forest Code, 2005).*

Thus, within the scope of this research we closely aligned the forest definition to the one of Forest Code Armenia, meaning, that the forests are defined as lands of more than 0.09 ha with the tree canopy cover of more than 30%. For this purpose, Hansen forest cover product for 2000 was adopted, as well as forest/non-forest (FNF) binary classification was performed using S2 product and Random Forest classifier (**Figure 12**).

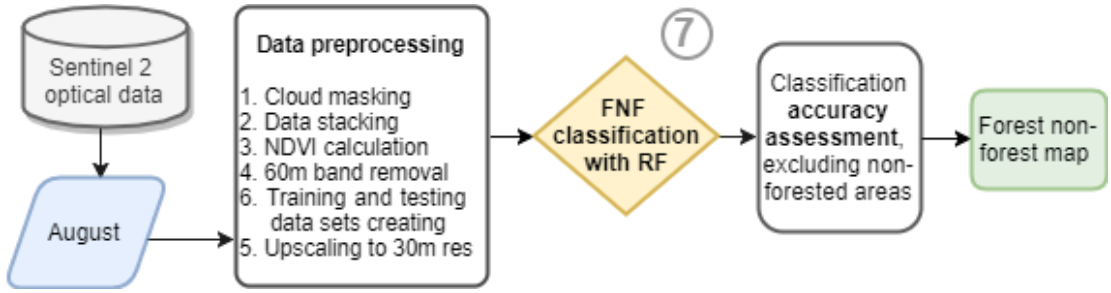



**Figure 12.** Flow chart for forest/non-forest classification steps

This FNF classification with a new data was necessary due to the literature indicating the scales of illegal felling (**Chapter 1.3**) and the Global Forest Watch online platform (Global Forest Watch), affirming that there are significant disturbance spots in the forests and there is a need for updating the Hansen map after 17 years. On the other hand, attempts to adjust Hansen Global map with simple threshold for forest non-forest delineation failed for this study area.

The training and testing datasets were generated and a binary classification was carried out in GEE cloud environment using JavaScript API and was performed in two steps. First, forest canopy cover threshold of 30% was setup using Hansen global forest cover map, as each pixel on this map represents the canopy cover percentage. This way we defined forest and non-forest areas for the year of 2000. That was particularly done for training dataset generation. For that purpose, "stratifiedSample" tool was used to generate randomly distributed points within a grid with a scale of 30m. 200 hundred points were generated this way: 100 points within the forest area and 100 points within the non-forest area.

**Validation confusion matrix**

| | | **Prediction** | | |
|---|---|---|---|---|
| | n =200 | **Positive** | **Negative** | Total |
| **Actual** | **Positive** | TP = 97 | FP = 3 | 100 |
| | **Negative** | FN = 4 | TN = 96 | 100 |
| | Total | 101 | 99 | 193 |

**Table 7**. Validation confusion matrix for forest non-forest classification

In the second step Random Forest classifier with 300 trees was trained using the S2 data from 2017 and the training dataset. FNF classification was driven and the fallowing accuracy parameters were assessed again in GEE environment (**Tables 7** ).

The RF performance on binary classification shows high **sensitivity**: 0.96 (proportion of actual positives that are correctly identified as such) and **specificity**: 0.97 (proportion of actual negatives that are correctly identified as such) using S2 data. Thus, the **overall accuracy** of classification is 0.965 and we use this classification result as a base map for our research to delineate only forested areas. **Kappa statistics** of the classification was 0.93. As the result, 86% (25184.4 ha) within the study area was classified as forest and 14% (4149.6 ha) was classified as non-forest (**Figure 13**). Majority voting filter was applied on the output raster map once.

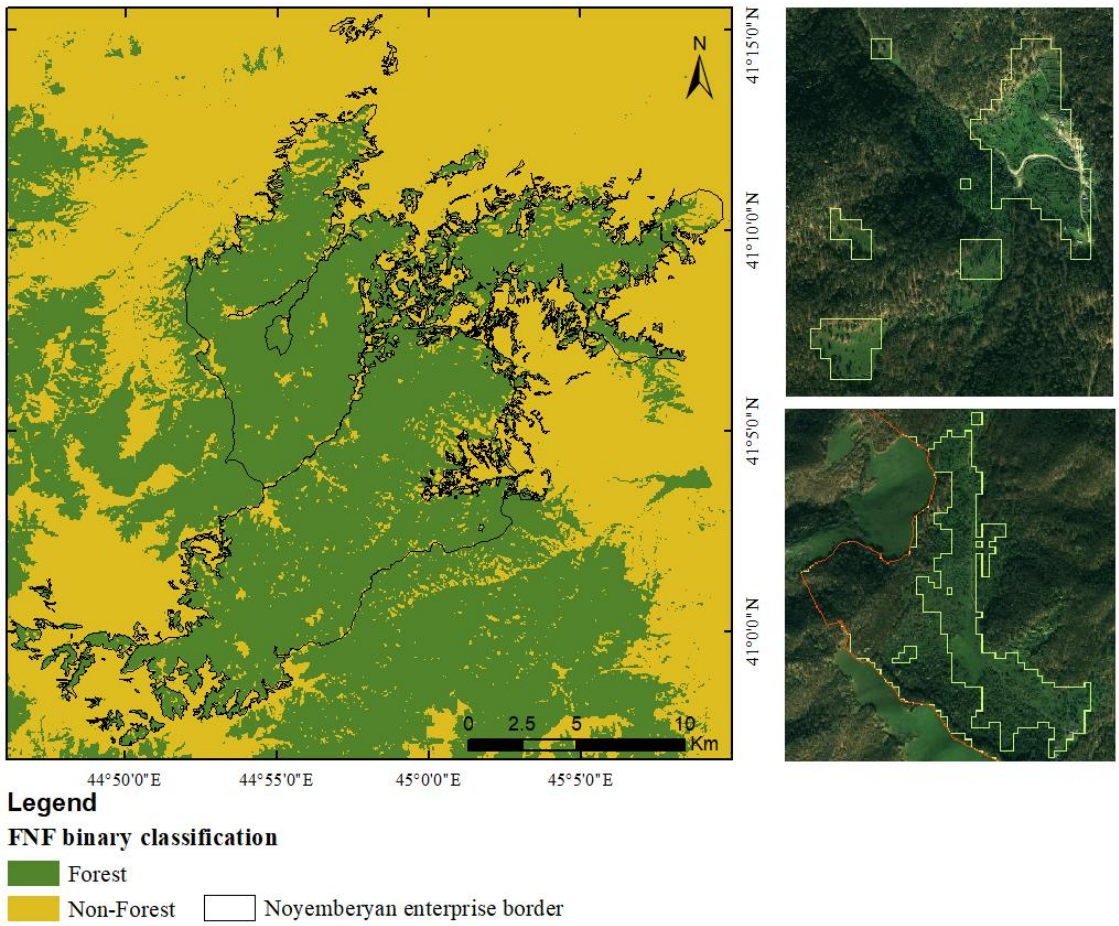



**Figure 13.** Map of FNF binary classification. Non-forested site examples on the snapshots on the right (green line shows the classified forest border, red line shows the forest enterprise border).

## 5.2 Results of Regression analysis

The regression analysis was conducted using AGB data (adjusted by scaling factor for 0.09 ha MMU) as a dependent variable and the rest of the inputs as independent variables (predictors). The sequence of the steps are following the general flow chart in the section of methodology (**Chapter 4.1**). Different combination of predictor variables were calculated and the regression model has been evaluated for multicollinearity in predictors (VIF test), normality (Shapiro-Wilk normality test) and autocorrelation (correlogram) in residuals for each time.

Regression analysis were carried out with R programming language and "Akaike criterion" function, which stands for multilinear stepwise regression. For the regression analysis total of 60 input variables were used (**Table 8**).

| Source | Number of variables | | Explanation Variable | Total |
|---|---|---|---|---|
| **Sentinel 1** | 12 | | VV monthly stack - 2017 | 32 |
| | 12 | | VH monthly stack - 2017 | |
| | 4 | GLCM | VV10 (Corr, Ent, Var, Savg) | |
| | 4 | | VH12 (Corr, Ent, Var, Savg) | |
| **Sentinel 2** | 3 | | NDVI, NDMI, RVI | 3 |
| **Alos Palsar** | 2 | | HH, HV - 2017 | 10 |
| | 4 | GLCM | HH (Corr, Ent, Var, Savg) | |
| | 4 | | HV (Corr, Ent, Var, Savg) | |
| **Ancillary data** | | | | |
| Aspect | 8 | | aspect categories | 15 |
| Slope | 1 | | slope values | |
| Forest type | 6 | | dominant tree type | |
| Total amount of variables | | | | 60 |

**Table 8.** Input variables for regression analysis

Initially, only 24 observations from S1 time series (VV and VH) were used for regression analysis in order to select the best scenes from S1 with higher importance in AGB prediction. This was done for two main purposes: First, to evaluate the preference of using S1 dense time series instead of choosing one scene for forest AGB calculations and second, to decrease the number of input variables for GLCM texture analysis for the further use. This was particularly important for keeping the number of input variables as less as possible, as the training dataset (field data) was not large.

From the perspective of the phytoclimatic seasonality in the forest in our study area, the correlation between field AGB and Sentinel 1 C-band stack time series were explored first using Pearson's correlation coefficient (**Figure 14**). This shows the correlation to be almost always higher for VH cross-polarization compared with VV co-polarization. Nevertheless, for both polarization modes it is lower than 0.5, and gets to its maximum for the stack averages of October and December (both for VH backscatter). The assumption is that C-band backscattering better represents the forest biomass during the leaves-off season, and when the soil is not frozen.

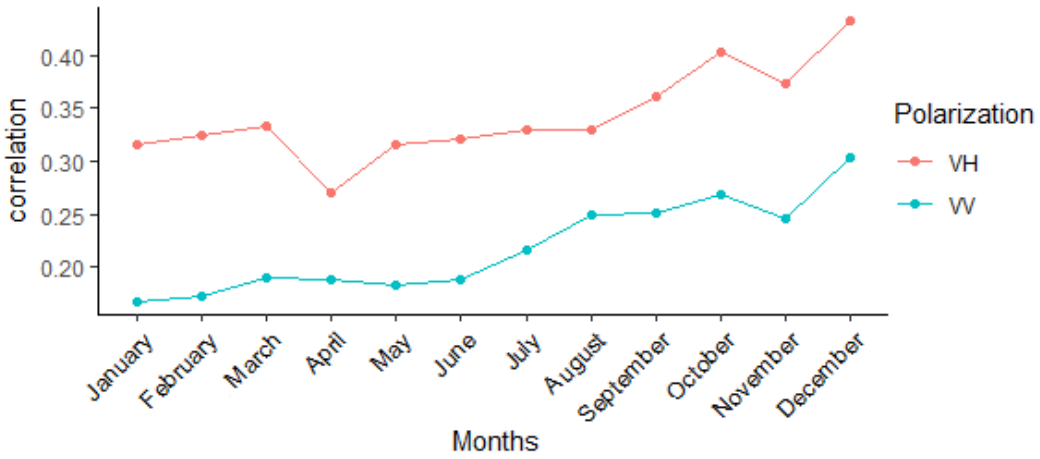



**Figure 14.** Pearson's correlation coefficient calculated between Sentinel 1 stack time series and AGB from the 79 plots and for each of the time series.

The regression analysis were repeated using the forest type information as a categorical variable. The results showed that among the S1 time series only the stacks of VH polarization from October and December are passing the minimum limit of p-value < 0.05. Also, both HH and HV polarization mosaics from Palsar were chosen by regression model to be important predictors for AGB. The forest type information was never selected as important, therefore was dropped from later analysis. Hence, we deduce that SAR backscattering is not sensitive to the tree type.

Those two stacks form S1 and HH, HV Palsar mosaics were chosen for GLCM texture measures and for further regression analysis. The GLCM analysis were done as it is described in the **Chapter 4.8**. Further, during this regression analysis, 2 samples with extreme biomass values for the study are (680 t $ha^{-1}$ and 500 t $ha^{-1}$), as well as 1 sample with unusual extreme pixel values on Palsar scenes were excluded from the further analysis, as they were causing major uncertainty in the prediction. As follows, 79 field sampling points were considered for the final regression analysis.

### 5.2.1 Model Comparison and Predictor Evaluation

Since the objective of this study is to evaluate the different origin data for forest AGB prediction, 4 regression models were proposed with different predictor combinations (**Table 10**). In this section we compared prediction models with a focus of detecting relative difference in the prediction accuracy. For this purpose, multiple linear regression analysis was carried out with different predictor fusion aiming to determine the best models using different type of data in different combinations. Accordingly, 4 models with the following combinations were developed: **Model 1** – only SAR data, **Model 2** – SAR+optical data, **Model 3** – SAR+ancillary data, **Model 4** – SAR+optical+ancillary data.

As many of the parameters are generated from varying decompositions of the same data or the same source, several of these metrics are expected to be highly correlated (**Appendix B**). High correlation amongst the predictors causes for model overfitting and biased $R^2$ value. Therefore, with the threshold of variance inflation factor maximum of 5 was set up. Adjusted $R^2$ is calculated. Further, Leave-One-Out cross-validation is carried out and $R^2$ as well as RMSE are calculated. As the RMSE represents the relative error for 0,09 ha area (MMU), later the absolute RMSE was calculated for the hectare and is provided in the **Table 9**.

Linear Model 1 consisting of S1 C-band and Palsar L-band SAR data and the GLCM texture analysis of those data yielded low prediction results: Adjusted $R^2$ equal to 0.46, LOO validation test gave $R^2$ equal to 0.38 and RMSE equal to 70 t per hectare. The best predictor variables were Palsar HH and HV polarization form the initial imagery, and the rest of the variables were different texture measures: Sum of Average (SAVG) GLCM for S1 stack of October, SAVG for Palsar HV. It is worth to mention, that none of the Sentinel initial images appeared in any of the models as important predictors.

Model 2 combined all the inputs from Model 1 and the 3 vegetation indices (NDVI, NDMI, RVI) from S2. Determination coefficient for LOO $R^2$ decreases insignificantly to 3.3 and the RMSE increases to 72.8 t $ha^{-1}$. The stepwise process discarded many of the variables that were important for the Model 1 and reduces the dimension of the predictors. Palsar HH and HV stay as the most important predictors together with Entropy measure of October stack from S1 data. NDVI variable was better suited for

AGB prediction amongst the vegetation indices with the less significance for the model.

| **Model** (sources) | **Predictors** | **VIF (<5)** | **p-value<0.05 (significance)** | | **Adj. $R^2$** | **LOO $R^2$** | **relative RMSE (t/0,09 ha)** | **RMSE (t/ha)** |
|---|---|---|---|---|---|---|---|---|
| **Model 1** SAR | Palsar HH | 2.26764 | 1.68E-05 | *** | 0.46 | **0.38** | 6.3 | **70** |
| | Palsar HV | 2.211967 | 1.18E-05 | *** | | | | |
| | S1 VH 10 savg | 1.913548 | 6.34E-05 | *** | | | | |
| | S1 VH 12 corr | 1.744905 | 0.03294 | * | | | | |
| | S1 VH 10 ent | 2.79781 | 0.03224 | * | | | | |
| | Palsar HH ent | 1.703185 | 0.00836 | ** | | | | |
| | Palsar HV savg | 2.345768 | 2.16E-05 | *** | | | | |
| | Palsar HH corr | 1.483653 | 0.04045 | * | | | | |
| **Model 2** SAR+ Optical | Palsar HH | 2.703908 | 3.24E-06 | *** | 0.41 | **0.33** | 65.6 | **72.8** |
| | Palsar HV | 2.21506 | 0.00045 | *** | | | | |
| | NDVI | 1.104523 | 0.022439 | * | | | | |
| | S1 VH 10 ent | 1.134397 | 0.000457 | *** | | | | |
| | Palsar HH savg | 2.323204 | 0.020287 | * | | | | |
| **Model 3** SAR+ Ancillary | Aspect (4,6,7) | 1.21354 | 5.00E-07 | *** | 0.65 | **0.6** | 50 | **57.7** |
| | Slope | 1.283596 | 7.15E-07 | *** | | | | |
| | Palsar HV | 2.930844 | 9.66E-06 | *** | | | | |
| | Palsar HH | 2.438455 | 0.002232 | *** | | | | |
| | Palsar HV savg | 1.721521 | 3.65E-10 | *** | | | | |
| | S1 VH 10 ent | 2.126821 | 9.23E-05 | *** | | | | |
| **Model 4** SAR+ Optical+ Ancillary | Aspect (4,6,7) | 1.216558 | 0.000842 | *** | 0.68 | **0.62** | 4.9 | **56.6** |
| | Slope | 1.290634 | 1.90E-07 | *** | | | | |
| | Palsar HV | 2.933416 | 4.20E-06 | *** | | | | |
| | Palsar HH | 2.459693 | 0.000854 | *** | | | | |
| | Palsar HV savg | 1.122131 | 0.000503 | *** | | | | |
| | S1 VH 10 ent | 1.736985 | 4.22E-10 | *** | | | | |
| | NDVI | 1.122131 | 0.019642 | * | | | | |

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05

**Table 9.** Evaluation of statistics of regression analysis for each model

The prediction accuracy changes enormously when the forest ancillary data (namely, aspect, slope and the forest type) were added to the previous data. Model 3 in **table 9** shows the overview of the statistics of regression analysis with all the inputs without data from optical sensor, and the Model 4 combines also the optical data. The LOO-validation $R^2$ for the Model 3 increases up to 0.6 and RMSE decreases to 57.7 t $ha^{-1}$. For the Model 4 the prediction improves very slightly with the implementation of

optical data: $R^2 = 0.62$ and RMSE = 56.6 t $ha^{-1}$. The only vegetation index selected by regression model as an important predictor is NDVI index for October. As in the previous models Palsar initial data with both polarizations was very significant predictor, as well as Sum Average texture measure of the Palsar HV polarization and the Entropy measure of the S1 stack for October.

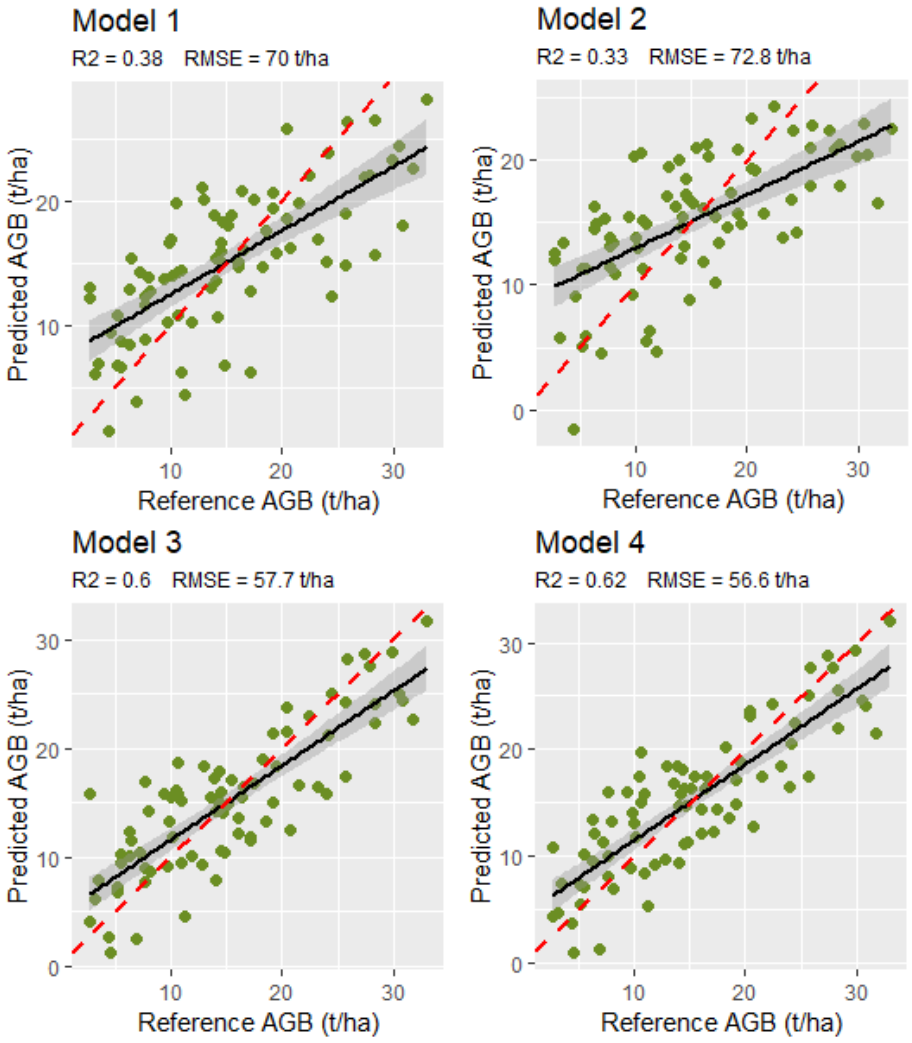



**Figure 15.** Linear regression of estimated above-ground biomass and reference above-ground biomass for 4 models (regression line in black, x = y line in red). Coefficient of determination $R^2$ and RMSE are presenting the results of LOO cross-validation

Backward stepwise regression for both Model 3 and 4 found very strong correlation between observed AGB data, SAR imagery and the topography data indicating that aspect and slope parameters appear in the models always as very important predictors. Aspect 4, 6 and 7 amongst the aspect categories were always in the list of most important predictors together with slope information. The methodology for implementation of ancillary data in order to establish better relationships between

satellite data and the forest stand characteristics in the mountainous areas brought huge impact on the AGB prediction meaning that the LOO-validation $R^2$ increased to 0.62 and the RMSE proportionally decreased to 56.6 t $ha^{-1}$.

It should be noted, that the significance of the predictors is certainly not between different models. For instance, Palsar HV polarization was the most significant for the models 1,2 and 4, but not for the model 3.

Scatterplot of forest AGB estimation versus the reference AGB (**Fig. 15**) displays similar distribution for the models 1 and 2 meaning that in both cases biomass is significantly overpredicted for small and close to average biomass, for the large biomass the models are underpredicting. Substantially improved prediction depict scatterplots for model 3 and 4. In both cases small biomass values tend to be slightly overpredicted and the larger ones are slightly underpredicted.

### 5.2.2 Model Comparison on Site Level

Due to the lack of further in-situ data on biomass for the study area we discuss the models in the site-level, i.e. we focus on comparing the relative biomass instead of ability to evaluate the exact biomass. As a final product of this research we used trained models for Forest AGB mapping over the study area. For selection of the final models for comparison and AGB map evaluation we took into consideration the model developed with the preliminary SAR data (model 1) and best final model with all the variables included (model 4). The **figure 16** presents the final AGB maps of those models, where we can see significant difference in both predicted biomass amount and the proportions of the biomass distribution. Model 1 as it shows the statistical analysis, predicted higher biomass with the mean values of 162 t $ha^{-1}$ for the whole area, while the mean of biomass for the model 4 is 140 t $ha^{-1}$ (calculated mean value for standing biomass is 167 t $ha^{-1}$ from 79 sampling pints).

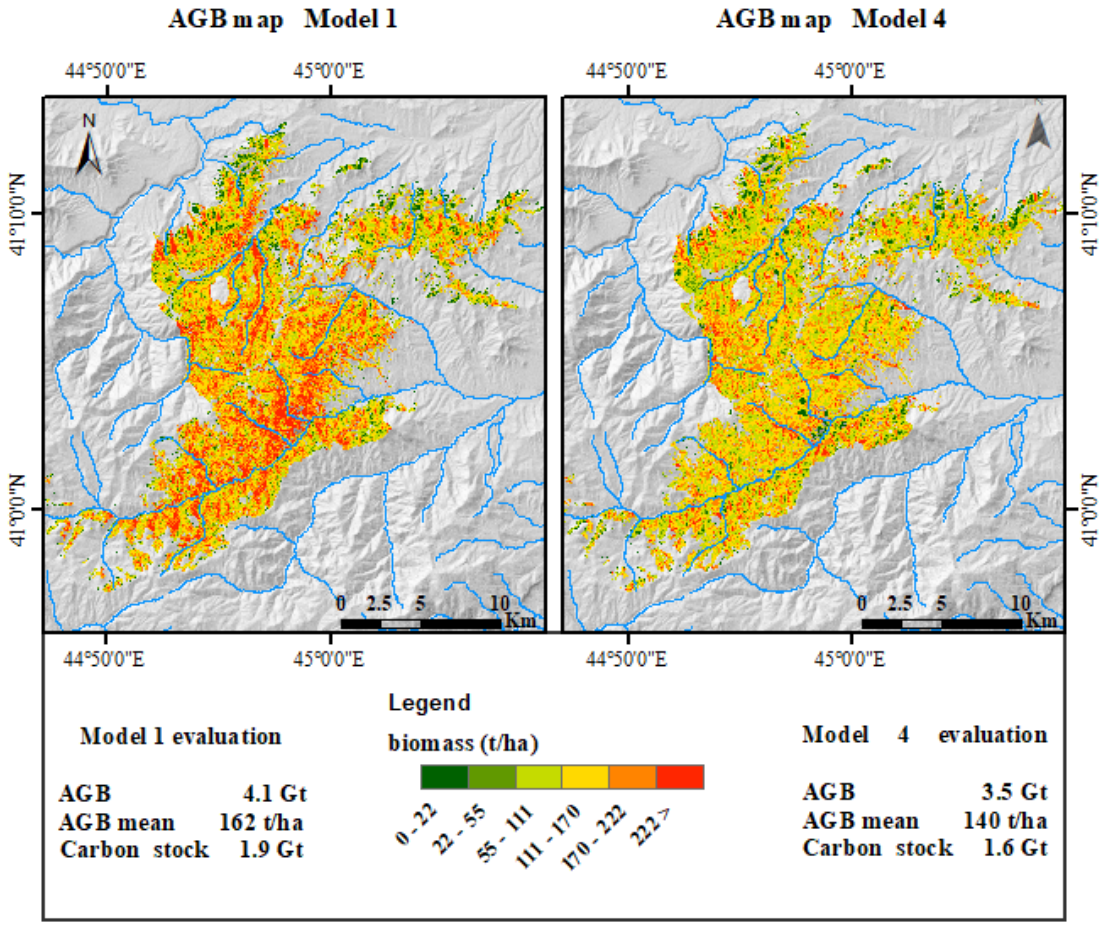



**Figure 16.** Forest AGB prediction maps for the study area with Model 1 and Model

Ancillary data were found to be useful for biomass prediction due to the fact, that it compensates the gain of SAR backscattering caused by the azimuthal terrain slope (model 4). In Contrast, the model 1 without ancillary data does overestimate the biomass because of the increased VH value, and visual interpretation shows notable repetition of topography over the whole study area. In all the complex topographic regions we can observe high biomass value altering respectively with the local variation of the slope and aspect. The **Figure 17** shows such example from a close view.

This means, that the use of SAR data and, especially, Sentinel 1 C-band SAR data in mountainous areas are very problematic. However, as the radar is sensitive to the dielectric characters and the moisture conditions of the target, S1 time series analysis were able to detect the most suitable month for biomass mapping and contribute to this study.

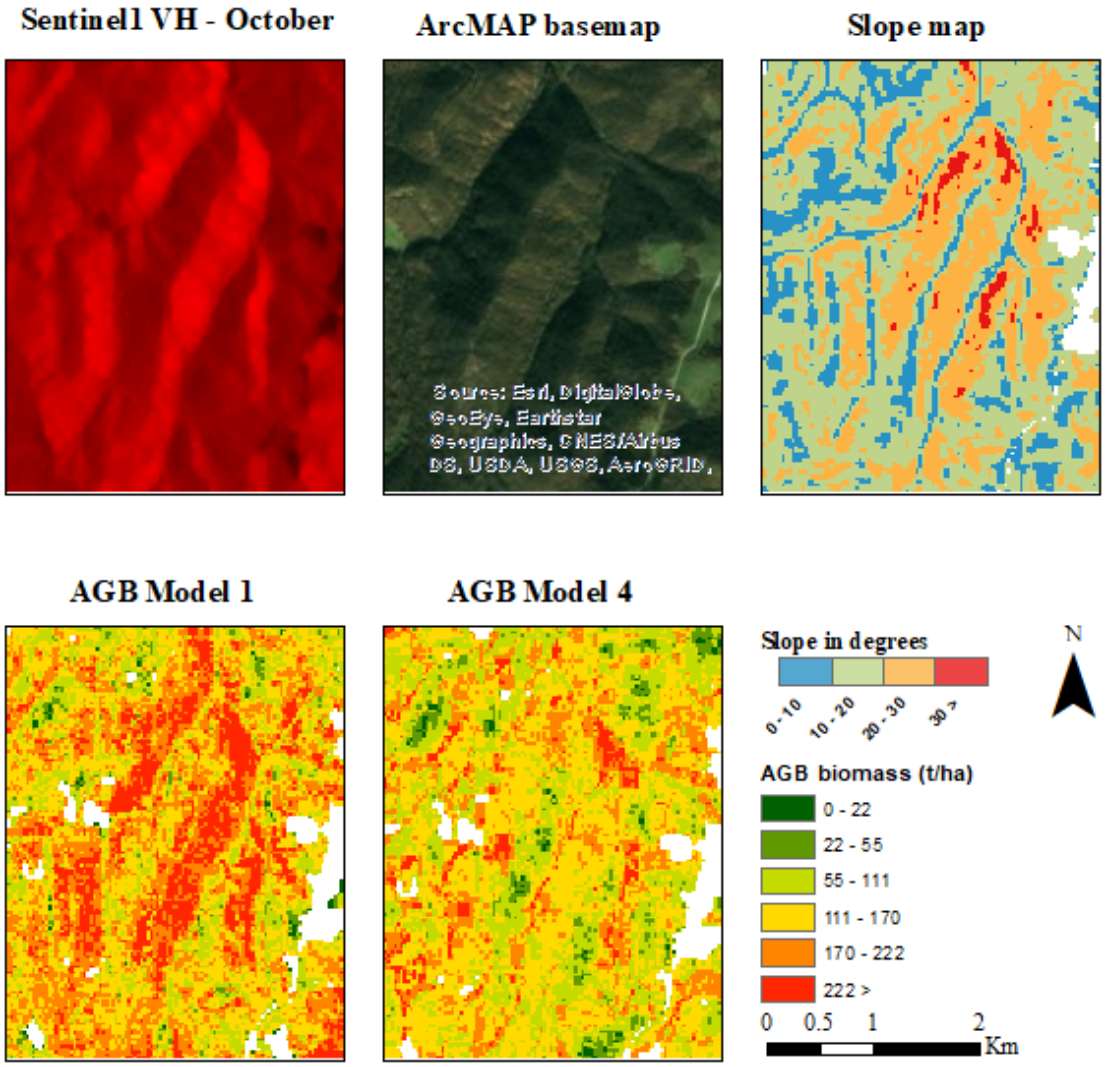



**Figure 17.** Snapshot of a study area with highlighted topographic effect on the AGB prediction. From left to right: topographic corrected Sentinel1 data stack of October, ESRI HR basemap, slope map from SRTM DEM, AGB prediction map with Model 1 ($R^2$ = 0.38, RMSE = 70 t $ha^{-1}$) and Model 4 ($R^2$ = 0.62, RMSE = 56.6 t $ha^{-1}$)

**Figure 18** shows the number of samples of reference AFB data in contrast of the predicted data by different models, when the sampling points are grouped in three different ranges based on AGB amount: first group includes the points with biomass less than 100 t $ha^{-1}$, second group combines the points with biomass between 100 and 200 t $ha^{-1}$ and the third group has the points bigger than 200 t $ha^{-1}$. The distribution of frequency of the prediction ranges depicts the main discrepancies between models as well as those difference from the reference data. The histogram of model 1 is affected by smaller frequency of first range in respect of second range, which is proportionally wider. The frequency for the biggest range (biomass > 200 t $ha^{-1}$) shows similar number of prediction and reference samples. The histogram of model 4

displays higher frequency for both small and middle ranges in respect to the reference much lower frequency for the range of >200 t ha$^{-1}$.

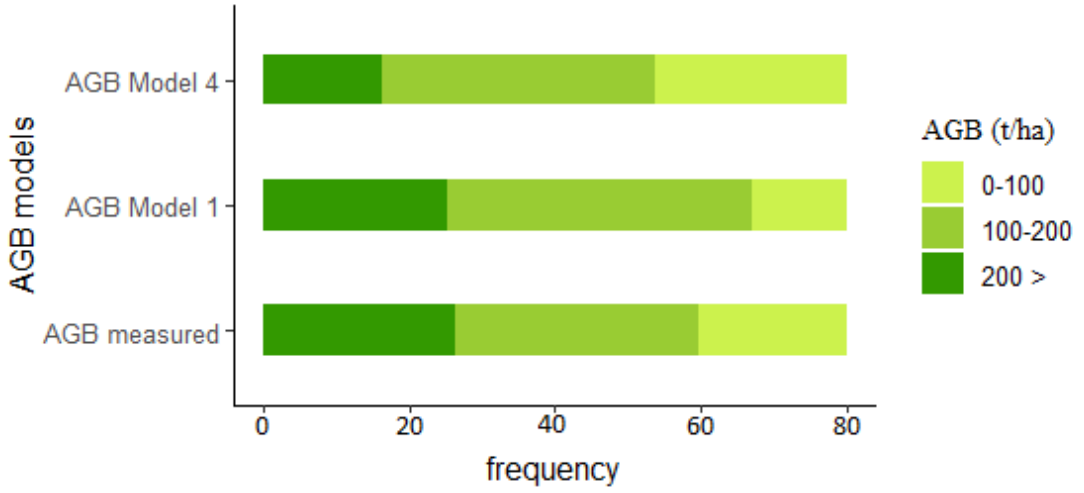



**Figure 18.** Measured and predicted above-ground biomass for the models 1 and 4. Frequency refers to the number of observations per range.

## 5.3 Discussion

The results consider MLR analysis using SAR backscattering combined with texture measures and optical data is useful for forest biomass modeling and has been widely excepted and extensively used in the last years. Although, those data have significant constrains over the mountainous areas and are recorded by many authors (Chen et al., 2018; Laurin et al., 2018; Vafaei et al., 2018), in this study we followed the advises of Chen et al, 2018, and Zhang et al., 2018, to implement forest ancillary data in the prediction model, which greatly improved the biomass prediction for the mountainous forests. Multiple linear stepwise regression model was selected for deeper understanding and better evaluation of multiple input variables and their role in the biomass prediction. Research showed that Sentinel 1 SAR dense time-series are useful not only for speckle reduction, but also multi-temporal analysis proved that the seasonal and forest phenological behavior greatly influence on the backscattering value. The correlation between SAR data and biomass value changes according to the seasonality as well as for different polarizations (**Fig.14**). The time period of late fall and the early winter is preferred for SAR analysis as the leaves are off and the ground is not frozen, so the shortwave C-band signal from the branches is more evident. Although, this regression analysis of time-series allowed to select the most relevant

month, the correlation of S1 data with biomass is still too low to be used for forest operational monitoring. Similar result and assumption is indicated in the work of Huang et al. ($R^2 = 0.28$) (Huang et al., 2018) and Berenguer ($R^2 = 0.40$) (Berenguer et al., 2018).

The prediction accuracy obtained only with C-band was low and there is need for integration of Palsar L-band data. This improved the model accuracy by increasing the adjusted $R^2$ to 0.46. Still the model was not stabile and the LOO validation drops $R^2$ to 0.38. This result is significantly lower compared with other studies' results which are using C and L-band radar data combination. Those two factors can explain this difference: First, they use the commercial Palsar L-band quad polarization data, which is greatly higher in resolution (fine resolution product - 8.7mx5.3m pixel spacing) and incidence angle is available, which allows to apply all the preferred pre-processing steps. Second, those articles make threshold either on SAR data to detect and exclude areas with foreshortening and layover (Laurin et al., 2018) or on the slope value, to exclude steep slopes to avoid from the same effect (Berninger et al., 2018).

For both SAR data cross-polarization backscattering there is higher correlation with biomass compared to co-polarization. For S1 it was always VH polarization selected by SWR model to be more important predictor. For Palsar data, even if both HH and HV polarizations were important in the predictions, HV cross-polarized backscattering was mostly superior over the HH co-polarization, although both of them were selected by the regression model as very important predictors and were not causing multicollinearity effect in the prediction. This result is observed in all the related papers.

This research is in agreement with the results of Berninger et al. and Chen et al. which show that the SAR image textures measures are significantly correlated to the AGB information and can strongly improve the prediction model parameters (Berninger et al., 2018; Chen et al., 2018). Within this study entropy and sum of average measures for S1 stack of October and correlation measure for S1 stack of December appeared to be more relative to the AGB than the initial backscatter intensity itself. For the Palsar data sum of average of HV and correlation and entropy measures were selected for different models as important predictors. Besides the high correlation between texture measures and AGB, the contribution of GLCM analysis to such researches is the fact, that the different measures from the same data are not highly correlated and can be

combined in the same model for regression analysis. Similarly, Huang uses 5 different measures form the same source for training the best model (Huang et al., 2018).

The correlation of optical data to forest AGB was very weak and only NDVI amongst the tested vegetation indices was selected by SWR with low importance. The insignificant contribution of optical data in the AGB prediction can be explained with the fact, that the average biomass for the forests of this study area was measured 148 $ha^{-1}$ while the observed saturation level for optical data in AGB prediction in the literature alters around 50 – 70 t $ha^{-1}$ biomass (Zhang et al., 2018). Although some other studies depict higher correlation between NDMI index and the biomass (literature review by Joshi et al., 2016), this analysis showed no similar results. It is possible that for NDMI calculation more humid season is preferred, while there was no cloudless image for the study area from March and June.

The use of ancillary data together with all the remotely sensed data brought a great impact on the regression model accuracy improving $R^2$ up to 0.62. This is the best result achieved within this research and is very important, as one of the hypotheses and the main innovation offered by this paper was the assumption that the gain or bias of the SAR backscattering value in mountainous areas cannot be corrected by the radiometric terrain correction methods, it might be mitigated by the prediction model when the forestry ancillary data on the sampling site is given. Amongst the ancillary data the slope and aspect information were always in a higher correlation with the biomass while combined with SAR data. Forest type information did not improve the biomass prediction in contrast of other studies (Laurin et al., 2018), which found it an important component for the prediction model. Even if the implementation of ancillary data greatly improves the prediction model performance as well as the visual observation of the product maps depict improvement in biomass distribution, more field data is needed to check the reliability of those models for implementing the radar remotely sensed data in carbon stock monitoring activities in mountainous forests.

Backward stepwise multiple linear regression technique considerably automated the best variable selection as well as isolated the multicollinearity effect. For instance, VIF test applied after SWR test did not detect multicollinearity in the variables.

Nevertheless, not all the studies use the same methods and measures for model diagnostics, which makes the result evaluation and comparison quite problematic. For

instance, many papers use only adjusted coefficient of determination ($R^2$) of regression models for the evaluation (Chen et al., 2018; Eckert, 2012; Huang et al., 2018), which is a results without model cross-validation, so we cannot directly compare our result with the ones from those works.

The saturation level for SAR backscattering as a general limitation, was hard to define as the backscattering value is suffering by the topography and cannot serve for such calibration. However, from the model 1 which uses SAR data only, we can observe level of 150-160 t $ha^{-1}$, after which the prediction line indicates mostly underprediction. This result is similar to the one from Berninger, 2018, even though he claims saturation level for L-band with texture measures are possible to improve up to 250 t $ha^{-1}$. The saturation level research can be explored in a more accurate way once more sampling data is available, which can make possible to separate only flat areas not affected by shadowing or foreshortening and do analysis without applying topographic information, even though diagram **(Fig. 18)** shows higher uncertainty for the low biomass prediction. This is likely because the measured trees are only the ones larger than 8 cm DBHOB, therefore smaller trees available in the sampling plots can have significant impact on the Sentinel 1 C-band (3.75–7.5 cm wavelength) backscattering causing for AGB overestimation which makes this range (<100 t $ha^{-1}$) the most influential. This trend of overestimation for samples with smaller biomass value is observed in all the similar researches (Berenguer et al., 2018; Berninger, et al., 2018; Huang et al., 2018; Laurin et al., 2018).

Even if very few of the reviewed papers perform new delineation of forest boundary, the importance of FNF classification performed within the frames of this paperwork is proven with two main points:

- FNF update shows significant (13%) non-forested areas within the forest enterprise, which should be taken into consideration while estimating the carbon stock for more realistic carbon budget evaluation.
- As the Hansen map is a global model, the FNF delineation based on that map was not adjustable for the study area. Therefore, the field sampling plots design generated based on the Hansen map can mislead the efficiency of the field work. As the result showed, 26 sampling plots out of 115 were located outside of the newly updated forest boundaries.

Thus, we conclude that the biomass data calculated by statistical methods using the allometric equations can be significantly affected by the reasons mentioned above. As this paper was aiming to imitate the forest carbon monitoring activities, and particularly, forest AGB measurements, we are planning to compare our results with the one from traditional NFI once it becomes available.

The final biomass map has 30 m spatial resolution, which is the first map having synoptic view on the spatial distribution of the biomass in NE forest of Armenia, yet with a comparable accuracy. The developed model is applicable for similar forests where in situ data is available. Another very important achievement is that this fine resolution map containing the proportional distribution of biomass can lead to more correct carbon estimation when used as source for carbon-related models. Therefore this study can facilitate the GEF/UNDP-REDD+ activities in NE forests in Armenia to achieve their objectives in better accuracy, as well as help to make forest carbon stock measurements and monitoring cost-effective. In order to calculate the carbon stock for the study area, the modeled forest stand biomass was reduced by a factor of 0.47 (UNDP-GEF 00091048, 2015). For the model 1 measured carbon stock in the study area was 4.1 Gt, and for the model 4 it was 3.5 Gt (**Fig. 16**). This data is yet to be compared with NFI results once it is available.

## 5.4 Limitations and Recommendations for Future Research

One of the major limitations for this study was the lack of flat areas that could be analysed separately from those areas with complex topography in order to be able to calibrate the true value of SAR backscattering for the forests in the study area. The next obstacle was the lack of sampling points. As seen in number of studies, the outcomes of such research are strongly dependant on the sampling data quantity and quality. As for the future, it is planned to collect total of 315 samples for this study area within the GEF-UNDP project. This amount of data itself is significantly bigger for model training and can make it possible to develop more sustainable prediction models and carry out research only on the relatively flat areas.

This study, on the other hand, made it obvious that complex topography puts high uncertainty on the study in the way, that forest structure and distribution changes quickly on very short distances, which is not desirable for SAR data. This is because it

is not very clear if the pixel on the SAR scene is fitting with the field sampling plot very well. Thus, the backscattering value can be presenting not the real biomass information of the sampling plot.

GLCM texture measures can be stronger or weaker correlated to forest biomass depending on the selected window size for the operation (Eckert, 2012). As for this research only one window size was adopted (4x4 kernel size), thus, we assume GLCM analysis to be not exhaustive and we consider potential possibility to improve prediction accuracy by adopting different window sizes for texture measures. On the other hand, texture analysis should be tested on the raw L-band radar data in order to keep the true speckle distribution and backscattering value statistics.

When field data becomes available, this methodology of biomass prediction can be applied per specific group of biomasses. We assume to have better correlation between SAR backscattering value and the biomass of under 150-160 t $ha^{-1}$ where the saturation level is observed.

As the relevance of the SAR data and its combination with other source data are well explored and the best input variables for forest AGB prediction are described, we propose application of non-parametric regressors for generating more accurate biomass map, as the review of the latest studies reveal their priority over parametric regressors (**Chapter 2.3**).

# CONCLUSION

Overall, this study shows the usefulness of SAR backscattering data on forest biomass mapping, which allows to have spatially explicit AGB distribution within the forest site. In this research, we propose a method based on the combination of multisource remotely sensed data, image texture measures and forest ancillary information for forest AGB estimation on the mountainous areas. For this purpose, only freely available data sources were used. The results infer the importance of multitemporal C-band data to avoid from backscattering saturation from moisture, also for multitemporal speckle filtering. However, there is a need of integration C-band with L-band data to reduce the uncertainty in the prediction. This study also showed that the GLCM texture characteristics of SAR data were the most relevant predictors for explaining the observed variability of AGB in the study area. The relationships between SAR data and stand characteristics, however, can be negatively influenced in areas of high topographic variation and make the AGB modelling truly problematic. The aspect and slope ancillary data was able to mitigate the topography effect on SAR data and ensured improved regression analysis with accuracy of $R^2 = 0.62$ and RMSE = 56.6 t $ha^{-1}$.

The output of this research will be provided to GEF/UNDP project in NE forests of Republic of Armenia and it can have an input in achieving their objectives, as well as supporting forest monitoring and carbon stock regulations. The achieved results need to be validated with more sophisticated ground-truth data in order to be implemented as cost-effective surrogate for NFI activities. This AGB estimation approach is adaptable and allows modeling of biomass in other mountainous forests with similar conditions.

Further, it is worth noting that working with SAR data demands prior knowledge and deep understanding of pre-processing steps and the physics behind them. It should also be said, that this study was carried out with freely available remotely sensed data, and more preferred commercial data availability can significantly increase the model performance accuracy.

# BIBLIOGRAPHIC REFERENCES


Aulard-Macler M., R. Barstow, D. R. (2011). Sentinel-1 Product Definition.

Balzter, H., Cole, B., Thiel, C., & Schmullius, C. (2015). Mapping CORINE Land Cover from Sentinel-1A SAR and SRTM Digital Elevation Model Data using Random Forests. Remote Sensing, 7(11), 14876. https://doi.org/10.3390/rs71114876

Berenguer, E., Oszwald, J., Id, I. T., Laurent, F., Hasan, A. F., & Sist, P. (2018). The Potential of Multisource Remote Sensing for Mapping the Biomass of a Degraded Amazonian Forest, 1–21. https://doi.org/10.3390/f9060303

Berninger, A., Lohberger, S., Stängel, M., & Siegert, F. (2018). SAR-Based Estimation of Above-Ground Biomass and Its Changes in Tropical Forests of Kalimantan, 2. https://doi.org/10.3390/rs10060831

Bouvet, A., Mermoz, S., Le Toan, T., Villard, L., Mathieu, R., Naidoo, L., & Asner, G. P. (2018). An above-ground biomass map of African savannahs and woodlands at 25 m resolution derived from ALOS PALSAR. Remote Sensing of Environment, 206(December 2017), 156–173. https://doi.org/10.1016/j.rse.2017.12.030

Bruniquel, J., & Lopes, A. (2010). Multi-variate optimal speckle reduction in SAR imagery, (November 2014), 37–41. https://doi.org/10.1080/014311697218962

CEOS. (2016). Interpretation Guide for ALOS PALSAR / ALOS-2 PALSAR-2 global 25 m mosaic data Table of Contents, 1.1, 1–15. Retrieved from http://www.gfoi.org/wp-content/uploads/2016/10/PALSAR-Interpretation-Guide-v1.1.pdf

Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., Eamus, D., … Yamakura, T. (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. Oecologia, 145(1), 87–99. https://doi.org/10.1007/s00442-005-0100-x

Chen, L., Ren, C., Zhang, B., Wang, Z., & Xi, Y. (2018). Estimation of Forest Above-Ground Biomass by Geographically Weighted Regression and Machine Learning with Sentinel Imagery, 1–20. https://doi.org/10.3390/f9100582

CRISP 2001. (n.d.). Principles of Remote Sensing - Centre for Remote Imaging, Sensing and Processing, CRISP. Retrieved October 19, 2018, from https://crisp.nus.edu.sg/~research/tutorial/sar_int.htm

de Badts, E. P. J. (2002). Global Land Cover 2000: Evaluation of the SPOT VEGETATION sensor for land use mapping, 49.

Dostálová, A., Wagner, W., Milenković, M., Hollaus, M., Dostálová, A., Wagner, W., … Wagner, W. (2018). Annual seasonality in Sentinel-1 signal for forest mapping and forest type classification. International Journal of Remote Sensing, 00(00), 1–23. https://doi.org/10.1080/01431161.2018.1479788

Dungan, J. L. (2002). Toward a Comprehensive View of Uncertainty in Remote Sensing Toward a Comprehensive View of Uncertainty in Remote Sensing Analysis Uncertainty in Remote Sensing Analysis. In Uncertainty in Remote Sensing and GIS. by Foody, M. Giles Atkinton, M. Peter (pp. 25–35). Department of Geography, University of Southampton, UK.

Eckert, S. (2012). Improved forest biomass and carbon estimations using texture measures from worldView-2 satellite data. Remote Sensing, 4(4), 810–829. https://doi.org/10.3390/rs4040810

Englhart, S., Keuck, V., & Siegert, F. (2012). Modeling aboveground biomass in tropical forests using multi-frequency SAR data-A comparison of methods. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5(1), 298–306. https://doi.org/10.1109/JSTARS.2011.2176720

ESAa. (2019). User Guides - Sentinel-1 SAR. Retrieved October 25, 2018, from https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/revisit-and-coverage

ESAc, Gatti, A., Naud, C., Castellani, C., & Carriero, F. (2018). Sentinel-2 Products Specification Document. Thales Alenia Space, 1–487.

Fassnacht, F. E., Hartig, F., Lati, H., Berger, C., Hernández, J., Corvalán, P., & Koch, B. (2014). Importance of sample size , data type and prediction method for remote sensing-based estimations of aboveground forest biomass, 154, 102–114. https://doi.org/10.1016/j.rse.2014.07.028

FMP. (2018). Noyemberyan Forest Enterprise: Forest Management Plan-2018.

Forest Code, A. (2005). Forest Code of The Republic of Armenia, 1–22.

G.F. De Grandi, M. Leysen, J.S. Lee, D. S. (2004). Radar Reflectivity Estimation Using Multiple SAR Scenes of the Same Target: Technique and Applications., (l), 1047–1050.

GEE API Tutorial. (2019). API tutorials | Google Earth Engine API. Retrieved January 5, 2019, from https://developers.google.com/earth-engine/tutorials

Ghazaryan, G., Dubovyk, O., Löw, F., Lavreniuk, M., Schellberg, J., & Kussul, N. (2018). A rule-based approach for crop identification using multi-temporal and multi-sensor phenological metrics. European Journal of Remote Sensing, 51(1), 511–524. https://doi.org/10.1080/22797254.2018.1455540

Gibbs, H. K., Brown, S., Niles, J. O., & Foley, J. A. (2007). Monitoring and estimating tropical forest carbon stocks : making REDD a reality. https://doi.org/10.1088/1748-9326/2/4/045023

Global Forest Watch. (n.d.). Retrieved January 20, 2019, from http://bit.ly/2R0Yg8f

Goetz, S., & Dubayah, R. (2011). Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change. Carbon Management, 2(3), 231–244. https://doi.org/10.4155/cmt.11.18

Haywood A. and Sayadyan H. and Hovhannisyan N. Implementing a simple design-based national forest carbon inventory in Armenia (Regular paper in preparation).2018.

Huang, X., Ziniti, B., Torbick, N., & Ducey, M. J. (2018). Assessment of Forest above Ground Biomass Estimation Using Multi-Temporal C-band Sentinel-1 and Polarimetric L-band PALSAR-2 Data. https://doi.org/10.3390/rs10091424

JAXA. (2018). Global 25 m Resolution PALSAR-2 / PALSAR Mosaic and Forest / Non-Forest Map ( FNF ): Dataset Description.

Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., … Ryan, C. M. (2016). A Review of the Application of Optical and Radar Remote Sensing Data Fusion to Land Use Mapping and Monitoring, 1–23. https://doi.org/10.3390/rs8010070

Kachamba, D. J., Ørka, H. O., Gobakken, T., Eid, T., & Mwase, W. (2016). Biomass estimation using 3D data from unmanned aerial vehicle imagery in a tropical woodland. Remote Sensing, 8(11). https://doi.org/10.3390/rs8110968

Kumar, K. K., Nagai, M., Witayangkurn, A., Kritiyutanant, K., & Nakamura, S. (2016). Above Ground Biomass Assessment from Combined Optical and SAR Remote Sensing Data in Surat Thani Province, Thailand. Journal of Geographic Information System, 08(04), 506–516. https://doi.org/10.4236/jgis.2016.84042

Kumar, L., Sinha, P., Taylor, S., & Alqurashi, A. F. (2015). Review of the use of remote sensing for biomass estimation to support renewable energy generation. Journal of Applied Remote Sensing, 9(1), 097696. https://doi.org/10.1117/1.JRS.9.097696

Laurin, G. V., Laurin, G. V., Balling, J., Corona, P., Mattioli, W., Papale, D., & Puletti, N. (2018). Above-ground biomass prediction by Sentinel-1 multitemporal data in central Italy with integration of ALOS2 and Sentinel-2 data, 12(1). https://doi.org/10.1117/1.JRS.12

Laurin, G. V., Pirotti, F., Callegari, M., Chen, Q., Cuozzo, G., Lingua, E., … Papale, D. (2016). Potential of ALOS2 and NDVI to Estimate Forest Above-Ground Biomass, and Comparison with Lidar-Derived Estimates. https://doi.org/10.3390/rs9010018

Liu, Z., Li, P., & Yang, J. (2017). Soil Moisture Retrieval and Spatiotemporal Pattern Analysis Using Sentinel-1 Data of Dahra, Senegal. Remote Sensing, 9(11), 1–18. https://doi.org/10.3390/rs9111197

Mermoz, S., Réjou-Méchain, M., Villard, L., Le Toan, T., Rossi, V., & Gourlet-Fleury, S. (2015). Decrease of L-band SAR backscatter with biomass of dense forests. Remote Sensing of Environment, 159, 307–317. https://doi.org/10.1016/j.rse.2014.12.019

MONGABAY. (2019). Forest data: Armenia Deforestation Rates and Related Forestry Figures. Retrieved January 7, 2019, from https://rainforests.mongabay.com/deforestation/archive/Armenia.htm?fbclid=IwAR0aRkzhqwoCCnb5FZocBp_MumAi2b5hcqH3U4-8KIa8k9jsB_YTXRyogJ0

Moreno-Sanchez, R., Sayadyan, H., Streeter, R., & Rozelle, J. (2007). The Armenian forests : threats to conservation and needs for sustainable management, (August). https://doi.org/10.2495/ECO070121

Navarro, A., Rolim, J., Miguel, I., Catalão, J., Silva, J., Painho, M., & Vekerdy, Z. (2016). Crop monitoring based on SPOT-5 Take-5 and sentinel-1A data for the estimation of crop water requirements. Remote Sensing, 8(6). https://doi.org/10.3390/rs8060525

Periasamy, S. (2018). Remote Sensing of Environment: Significance of dual polarimetric synthetic aperture radar in biomass retrieval : An attempt on Sentinel-1. Remote Sensing of Environment, 217(August), 537–549. https://doi.org/10.1016/j.rse.2018.09.003

Pohl C., G. J. (2017). Remote sensing image fusion: A practical guide. https://doi.org/10.1117/12.226824

Potin, P., Rosich, B., Miranda, N., & Grimont, P. (2016). Sentinel-1 Mission Status. Procedia - Procedia Computer Science, 100, 1297–1304. https://doi.org/10.1016/j.procs.2016.09.245

Pottier E, Lee, J., & Ferro-Famil, L. (2004). ADVANCED CONCEPTS IN POLARIMETRY – PART 1 (Polarimetric Target Description, Speckle filtering and Decomposition Theorems). Nato Otan, 1(1), 1–29. https://doi.org/10.1.1.214.8632

R Core Team. (2018). R: A Language and Environment for Statistical Computing. Retrieved February 22, 2019, from: https://www.r-project.org/

Reiche, J., Hamunyela, E., Verbesselt, J., Hoekman, D., & Herold, M. (2017). Remote Sensing of Environment Improving near-real time deforestation monitoring in tropical dry forests by combining dense Sentinel-1 time series with Landsat and ALOS-2 PALSAR-2. Remote Sensing of Environment, 204(April), 0–1. https://doi.org/10.1016/j.rse.2017.10.034

Reiche, J., Lucas, R., Mitchell, A. L., Verbesselt, J., Hoekman, D. H., Haarpaintner, J., … Herold, M. (2016). Combining satellite data for better tropical forest monitoring. Nature Publishing Group, 6(2), 120–122. https://doi.org/10.1038/nclimate2919

Reiche, J., Verhoeven, R., Verbesselt, J., Wielaard, N., Herold, M., Hamunyela, E., … Herold, M. (2018). Characterizing Tropical Forest Cover Loss Using Dense Sentinel-1 Data and Active Fire Alerts. REMOTE SENSING, 10(5), 1–18. https://doi.org/10.3390/rs10050777

Rio+20. (2012). National Assessment Report: Armenia.

Rüetschi, M., Schaepman, M. E., Small, D., Rueetschi, M., Schaepman, M. E., Small, D., Small, D. (2018). Using Multitemporal Sentinel-1 C-band Backscatter to Monitor Phenology and Classify Deciduous and Coniferous Forests in Northern Switzerland. Remote Sensing, 10(1), 1–N.PAG. https://doi.org/10.3390/rs10010055

Santi, E., Paloscia, S., Pettinato, S., Fontanelli, G., Mura, M., Zolli, C., … Chirici, G. (2017). The potential of multifrequency SAR images for estimating forest biomass in Mediterranean areas. Remote Sensing of Environment, 200(July), 63–73. https://doi.org/10.1016/j.rse.2017.07.038

Santoro, M., & Cartus, O. (2018). Research pathways of forest above-ground biomass estimation based on SAR backscatter and interferometric SAR observations. Remote Sensing, 10(4). https://doi.org/10.3390/rs10040608

Sayadyan H.Y., Avetisyan G.D. and Vardanyan E.A. Issues related to the improvement of forest landscape management in Tavush region. Biolog. Journal of Armenia, 4(69), 2017, pp.108-111.

Shahrokhzadeh, U., Sohrabi, H., & Copenheaver, C. A. (2015). Aboveground biomass and leaf area equations for three common tree species of Hyrcanian temperate forests in northern Iran. Botany, 93(10), 663–670. https://doi.org/10.1139/cjb-2015-0078

Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraishi, T., Thapa, R., & Lucas, R. (2014). New global forest/non-forest maps from ALOS PALSAR data (2007-2010). Remote Sensing of Environment, 155, 13–31. https://doi.org/10.1016/j.rse.2014.04.014

Sinha, S., Jeganathan, C., Sharma, L. K., & Nathawat, M. S. (2015). A review of radar remote sensing for biomass estimation.https://doi.org/10.1007/s13762-015-0750-0

Skakun, S., Kussul, N., Shelestov, A. Y., Lavreniuk, M., & Kussul, O. (2015). Efficiency Assessment of Multitemporal C-Band Radarsat-2 Intensity and Landsat-8 Surface Reflectance Satellite Imagery for Crop Classification in Ukraine. IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing. https://doi.org/10.1109/JSTARS.2015.2454297

Small, D. (2011). Flattening Gamma: Radiometric Terrain Correction for SAR Imagery. Strategy and Leadership, 38(1), 11–16. https://doi.org/10.1108/10878571011009831

Soenen, S. A., Peddle, D. R., Hall, R. J., Coburn, C. A., & Hall, F. G. (2010). Estimating aboveground forest biomass from canopy reflectance model inversion in mountainous terrain. Remote Sensing of Environment, 114(7), 1325–1337. https://doi.org/10.1016/j.rse.2009.12.012

Thapa, R. B., Watanabe, M., Shimada, M., & Motohka, T. (2016). Examining High-Resolution PiSAR-L2 Textures for Estimating Tropical Forest Carbon Stocks. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(7), 3202–3209. https://doi.org/10.1109/JSTARS.2016.2528262

Tian, B., Li, Z., Xu, J., Fu, S., & Liu, J. (2014). Mapping mountain meadow with high resolution and polarimetric SAR data. IOP Conference Series: Earth and Environmental Science, 17(1). https://doi.org/10.1088/1755-1315/17/1/012076

Timothy, D., Onisimo, M., & Riyad, I. (2016). Quantifying aboveground biomass in African environments: A review of the trade-offs between sensor estimation accuracy and costs, 57(3), 393–405.

Torres, R., Lokas, S., Cosimo, G., Geudtner, D., & Bibby, D. (2017). Sentinel 1 Evolution: Sentinel-1c And -1d Models, 5549–5550.

Treuhaft, R., Id, Y. L., Gonçalves, F., Keller, M., Roberto, J., Id, M. N., & Almeida, A. (2017). Tropical-Forest Structure and Biomass Dynamics from TanDEM-X Radar Interferometry, 1–28. https://doi.org/10.3390/f8080277

UNDP-GEF 00091048, U.-G. (2015). Mainstreaming Sustainable Land and Forest Management in Mountain Landscapes of North-eastern Armenia, (i).

Vafaei, S., Soosani, J., Adeli, K., Fadaei, H., Naghavi, H., Pham, T. D., & Bui, D. T. (2018). Improving accuracy estimation of Forest Aboveground Biomass based on

incorporation of ALOS-2 PALSAR-2 and Sentinel-2A imagery and machine learning: A case study of the Hyrcanian forest area (Iran). Remote Sensing, 10(2). https://doi.org/10.3390/rs10020172

Whelen, T., & Siqueira, P. (2018). Time-series classification of Sentinel-1 agricultural data over North Dakota. Remote Sensing Letters, 9(5), 411. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=edb&AN=128358742&site=eds-live&authtype=ip,uid

Yu, Y., & Saatchi, S. (2016). Sensitivity of L-band SAR backscatter to aboveground biomass of global forests. Remote Sensing, 8(6). https://doi.org/10.3390/rs8060522

Zhang, H., Wang, C., Zhu, J., Fu, H., Xie, Q., & Shen, P. (2018). Forest above-ground biomass estimation using single-baseline polarization coherence tomography with P-Band PolInSAR data. Forests, 9(4). https://doi.org/10.3390/f9040163

## APPENDIX A – LINKS TO THE CODES

Link address to FNF classification code in GEE API platform:

https://code.earthengine.google.com/8e7a4e2076a6c772727eae5b57e3ef53

Link address to Sentinel 1 pre-processing code in GEE API platform:

https://code.earthengine.google.com/2fa5e2ce872819e3941e6bb2e0717979

Link address to Alos Palsar pre-processing code in GEE API platform:
https://code.earthengine.google.com/1c642faaff184c442a723f7b33d5ac26

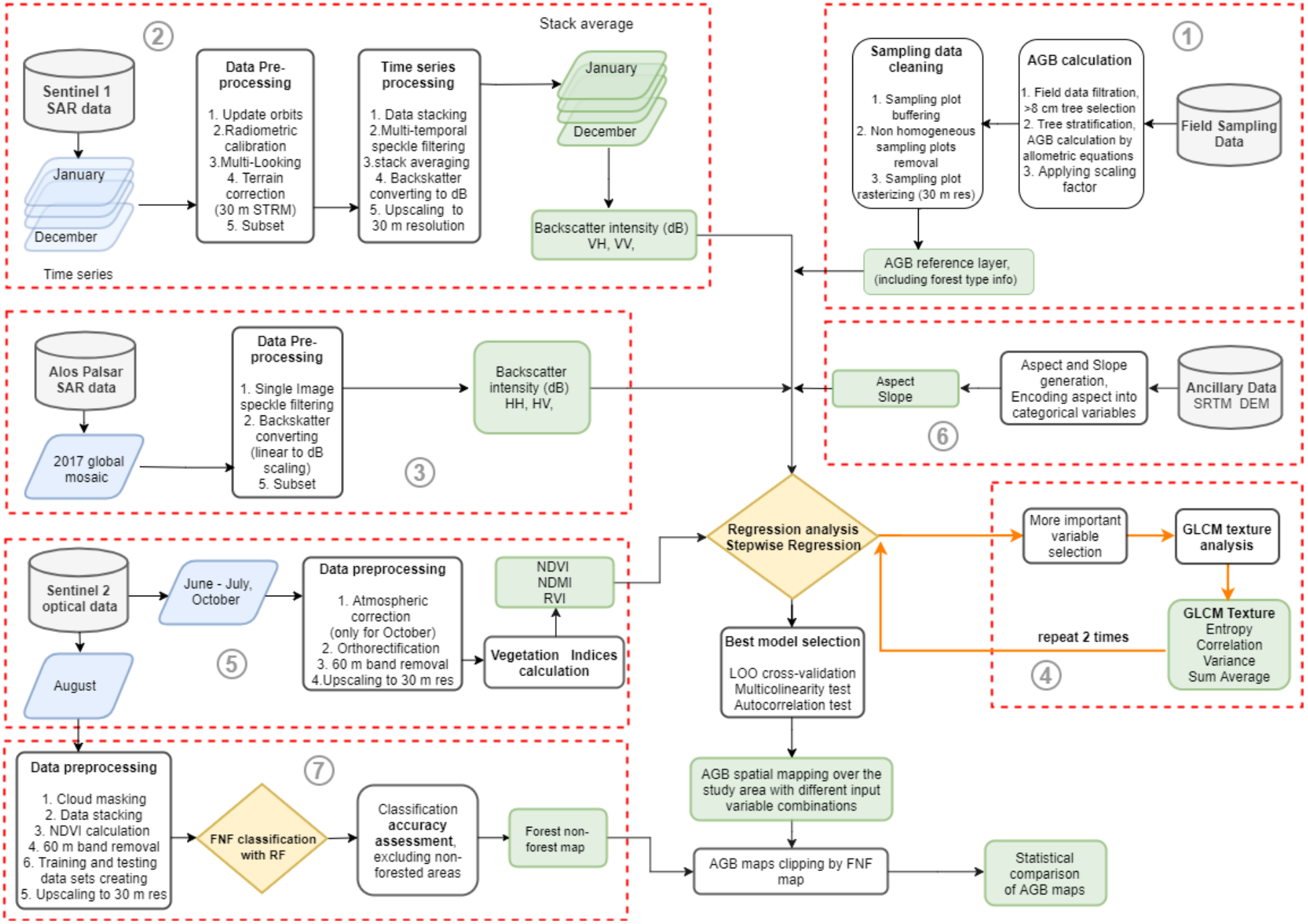
Link address to vegetation indices calculations code in GEE API platform:

https://code.earthengine.google.com/442c6d44b45b51b7fc9704852246fc27

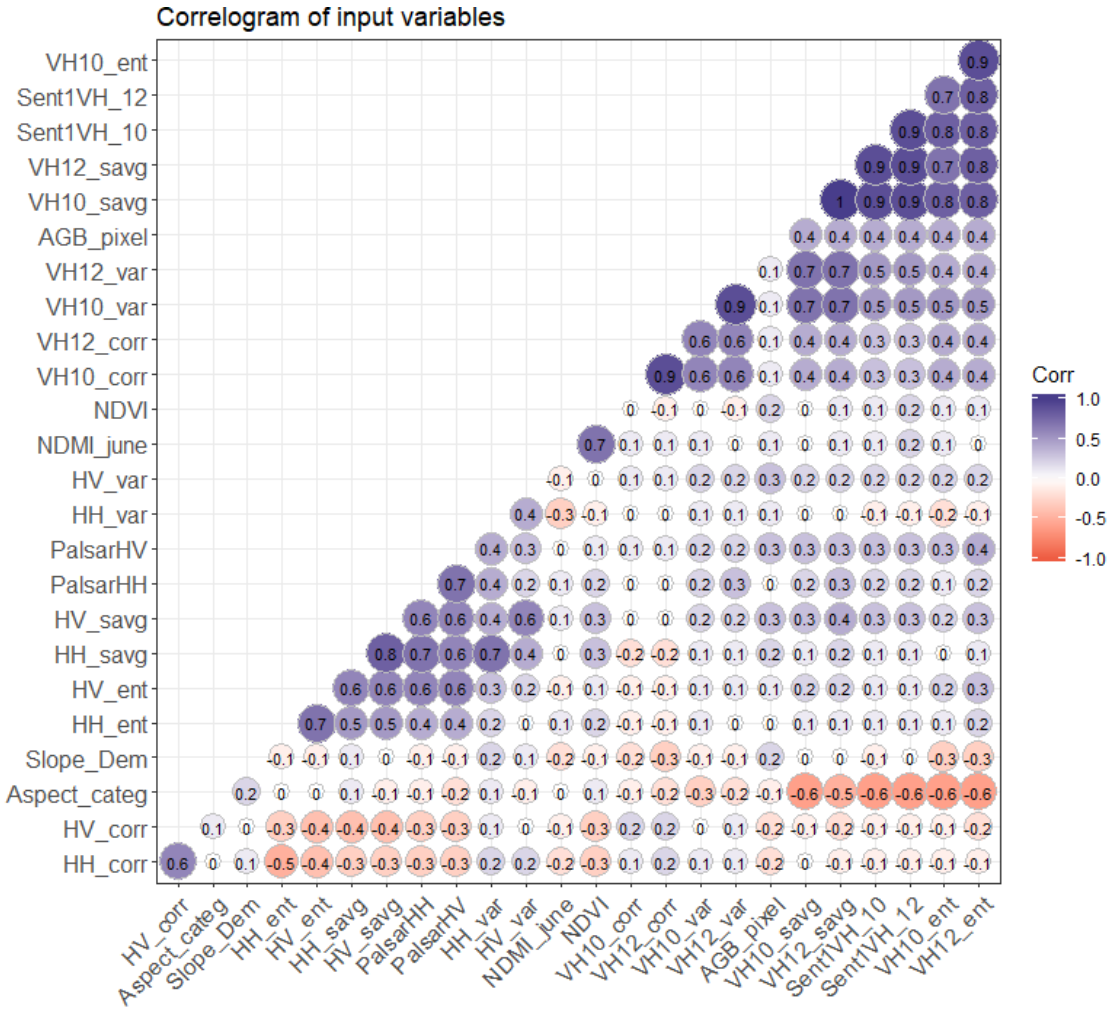
Link address to regression analysis code in Git-Hub. Codes are in R programming language:

https://github.com/armkhudinyan/SWR_analysis

## APPENDIX B - FLOW CHART FOR OVERALL METHODOLOGY


2
Sentinel 1 SAR data
January
December
Time series
Data Pre-processing
1. Update orbits
2.Radiometric calibration
3.Multi-Looking
4. Terrain correction (30 m STRM)
5. Subset
Time series processing
1. Data stacking
2.Multi-temporal speckle filtering
3.stack averaging
4. Backskatter converting to dB
5. Upscaling to 30 m resolution
Stack average
January
December
Backscatter intensity (dB) VH, VV,
1
Sampling data cleaning
1. Sampling plot buffering
2. Non homogeneous sampling plots removal
3. Sampling plot rasterizing (30 m res)
AGB calculation
1. Field data filtration, >8 cm tree selection
2. Tree stratification, AGB calculation by allometric equations
3. Applying scaling factor
Field Sampling Data
AGB reference layer, (including forest type info)
Alos Palsar SAR data
2017 global mosaic
Data Pre-processing
1. Single Image speckle filtering
2. Backskatter converting (linear to dB scaling)
5. Subset
3
Backscatter intensity (dB) HH, HV,
Aspect
Slope
Aspect and Slope generation, Encoding aspect into categorical variables
Ancillary Data SRTM DEM
6
Sentinel 2 optical data
August
June - July, October
5
Data preprocessing
1. Atmospheric correction (only for October)
2. Orthorectification
3. 60 m band removal
4.Upscaling to 30 m res
Vegetation Indices calculation
NDVI
NDMI
RVI
Regression analysis Stepwise Regression
Best model selection
LOO cross-validation
Multicolinearity test
Autocorrelation test
More important variable selection
GLCM texture analysis
GLCM Texture
Entropy
Correlation
Variance
Sum Average
repeat 2 times
4
7
Data preprocessing
1. Cloud masking
2. Data stacking
3. NDVI calculation
4. 60 m band removal
6. Training and testing data sets creating
5. Upscaling to 30 m res
FNF classification with RF
Classification accuracy assessment, excluding non-forested areas
Forest non-forest map
AGB spatial mapping over the study area with different input variable combinations
AGB maps clipping by FNF map
Statistical comparison of AGB maps

## APPENDIX C – CORRELATION MATRIX




Correlation matrix of the final input variables for regression analysis

# Masters Program in **Geospatial Technologies**

Supported by:
Education and Culture
ERASMUS MUNDUS