
Master's thesis in Geography
Geoinformatics

# ASSESSING *AGAVE SISALANA* BIOMASS FROM LEAF TO PLANTATION LEVEL USING FIELD MEASUREMENTS AND MULTISPECTRAL SATELLITE IMAGERY

Ilja Vuorinne

2020

Supervisors:
Janne Heiskanen
Petri Pellikka

Master's Programme in Geography
Faculty of Science

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

| Tiedekunta – Fakultet – Faculty | Osasto – Institution – Department |
|---|---|
| Faculty of Science | Department of Geosciences and Geography |

Tekijä – Författare – Author

Ilja Vuorinne

Tutkielman otsikko – Avhandlings titel – Title of thesis

Assessing *Agave sisalana* biomass from leaf to plantation level using field measurements and multispectral satellite imagery

Koulutusohjelma ja opintosuunta – Utbildningsprogram och studieinriktning – Programme and study track

Master's programme in geography, Geoinformatics

| Tutkielman taso – Avhandlings nivå – Level of the thesis | Aika – Datum – Date | Sivumäärä – Sidoantal – Number of pages |
|---|---|---|
| Master's thesis | June 2020 | 56 + appendices |

Tiivistelmä – Referat – Abstract

Biomass is an important parameter for crop monitoring and management, as well as for assessing carbon cycle. In the field, allometric models can be used for non-destructive biomass assessment, whereas remote sensing is a convenient method for upscaling the biomass estimations over large areas. This study assessed the dry leaf biomass of *Agave sisalana* (sisal), a perennial crop whose leaves are grown for fibre and biofuel production in tropical and subtropical regions. First, an allometric model was developed for predicting the leaf biomass. Then, Sentinel-2 multispectral satellite imagery was used to model the leaf biomass at 8851 ha plantation in South-Eastern Kenya.

For the allometric model 38 leaves were sampled and measured. Plant height and leaf maximum diameter were combined into a volume approximation and the relation to biomass was formalised with linear regression. A strong log-log linear relation was found and leave-one-out cross-validation for the model showed good prediction accuracy ($R^2$ = 0.96, RMSE = 7.69g). The model was used to predict biomass for 58 field plots, which constituted a sample for modelling the biomass with Sentinel-2 data. Generalised additive models were then used to explore how well biomass was explained by various spectral vegetation indices (VIs). The highest performance ($D^2$ = 74%, RMSE = 4.96 Mg/ha) was achieved with VIs based on the red-edge ($R_{740}$ and $R_{783}$), near-infrared ($R_{865}$) and green ($R_{560}$) spectral bands. Highly heterogeneous growing conditions, mainly variation in the understory vegetation seemed to be the main factor limiting the model performance. The best performing VI ($R_{740}/R_{783}$) was used to predict the biomass at plantation level. The leaf biomass ranged from 0 to 45.1 Mg/ha, with mean at 9.9 Mg/ha.

This research resulted a newly established allometric equation that can be used as an accurate tool for predicting the leaf biomass of sisal. Further research is required to account for other parts of the plant, such as the stem and the roots. The biomass-VI modelling results showed that multispectral data is suitable for assessing sisal leaf biomass over large areas, but the heterogeneity of the understory vegetation limits the model performance. Future research should address this by investigating the background effects of understory and by looking into complementary data sources. The carbon stored in the leaf biomass at the plantation corresponds to that in the woody aboveground biomass of natural bushlands in the area. Future research is needed on soil carbon sequestration and soil and plant carbon fluxes, to fully understand the carbon cycle at sisal plantation.

Avainsanat – Nyckelord – Keywords

agave sisalana, biomass, allometry, linear regression, remote sensing, vegetation indices, sentinel-2, generalized additive model

Säilytyspaikka – Förvaringställe – Where deposited

University of Helsinki electronic theses library E-thesis/HELDA

Muita tietoja – Övriga uppgifter – Additional information

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

| Tiedekunta – Fakultet – Faculty | Osasto – Institution – Department |
|---|---|
| Matemaattis-luonnontieteellinen tiedekunta | Geotieteiden ja maantieteen laitos |

Tekijä – Författare – Author

Ilja Vuorinne

Tutkielman otsikko – Avhandlings titel – Title of thesis

Assessing *Agave sisalana* biomass from leaf to plantation level using field measurements and multispectral satellite imagery

Koulutusohjelma ja opintosuunta – Utbildningsprogram och studieinriktning – Programme and study track

Maantieteen koulutusohjelma, Geoinformatiikka

| Tutkielman taso – Avhandlings nivå – Level of the thesis | Aika – Datum – Date | Sivumäärä – Sidoantal – Number of pages |
|---|---|---|
| Pro gradu -tutkielma | Kesäkuu 2020 | 56 + liitteet |

Tiivistelmä – Referat – Abstract

Biomassa, eli kasviaineksen määrä, on tärkeä muuttuja viljelykasvien kasvun seurannassa sekä arvioitaessa hiilen kiertoa. Kenttätöissä biomassaa voidaan arvioida kasveja vahingoittamatta hyödyntämällä allometrisia malleja. Suuremmassa mittakaavassa biomassaa voidaan kartoittaa kaukokartoitusmenetelmillä. Tässä tutkimuksessa arvioitiin *Agave sisalanan* eli sisalin lehtien kuivaa biomassaa. Sisal on trooppisilla ja subtrooppisilla alueilla viljeltävä monivuotinen kasvi, jonka lehdistä tuotetaan kuitua ja biopolttoainetta. Lehtibiomassan arvioimiseksi luotiin ensin allometrinen malli, minkä jälkeen biomassa mallinnettiin 8851 hehtaarin plantaasille Kaakkois-Keniassa käyttämällä Sentinel-2 multispektraalista satellittikuva-aineistoa.

Allometrista mallia varten kerättiin 38:n lehden otos. Kasvin korkeuden ja lehden suurimman ympärysmitan avulla muodostettiin tilavuusarvio, jonka yhteyttä biomassaan mallinnettiin lineaarisella regressiolla. Muuttujien välille löytyi vahva log-log lineaarinen yhteys ja ristiinvalidointi osoitti, että mallin ennusteet ovat tarkkoja ($R^2 = 0.96$, RMSE = 7.69g). Mallin avulla ennustettiin lehtibiomassa 58:lle koealalle, jotka muodostivat otoksen biomassan mallinnukseen Sentinel-2 kuvalla. Mallinnuksessa käytettiin yleistettyjä additiivisia malleja, joiden avulla tutkittiin lukuisten spektraalisten kasvillisuusindeksien yhteyttä biomassaan. Parhaaksi osoittautuivat indeksit, jotka laskettiin hyödyntämällä vihreää ja lähi-infrapunakanavaa, sekä ns. ”red-edge”-kanavia ($D^2 = 74\%$, RMSE = 4.96 Mg/ha). Keskeisin mallin selitysastetta heikentävä tekijä vaikutti olevan suuresti vaihteleva aluskasvillisuuden määrä. Hyödyntämällä parhaaksi todettua kasvillisuusindeksiä lehtibiomassa mallinnettiin koko plantaasin peltoalalle. Biomassa vaihteli 0 ja 45.1 Mg/ha välillä, keskiarvon ollessa 9.9 Mg/ha.

Tämän tutkimuksen tuloksena syntyi allometrinen malli, jota voidaan käyttää sisalin lehtibiomassan arviointiin. Jatkotutkimuksissa tulisi ottaa huomioon myös kasvin muut osat, kuten varsi ja juuret. Biomassan mallinnus multispektraalisilla kasvillisuusindekseillä osoitti menetelmän toimivuuden sisalin biomassan kartoituksessa, mutta vaihtelevan aluskasvillisuuden todettiin heikentävän mallin suorituskykyä. Aluskasvillisuuden vaikutusta ja täydentäviä aineistolähteitä tulisi tutkia tulevaisuudessa. Plantaasin lehtibiomassan, ja näin ollen maanpäälle sitoutuneen hiilen määrä, on saman suuruinen, kuin alueen luonnollisella pensassavannilla. Sisal-plantaasin hiilen kierron kokonaisvaltainen ymmärtäminen vaatii kuitenkin lisätietoa kasvien ja maaperän hiilivuosta sekä maaperän hiilensitomisesta.

Avainsanat – Nyckelord – Keywords

agave sisalana, biomassa, allometria, lineaarinen regressio, kaukokartoitus, kasvillisuusindeksit, sentinel-2

Säilytyspaikka – Förvaringställe – Where deposited

University of Helsinki electronic theses library E-thesis/HELDA

Muita tietoja – Övriga uppgifter – Additional information

Table of Contents

## Abbreviations

| | |
|---|---|
| AGB | Aboveground biomass |
| B | Blue |
| CI | Chlorophyll index |
| $CH_4$ | Methane |
| $C_2O$ | Carbon dioxide |
| G | Green |
| ESA | European Space Agency |
| FAO | Food and Agriculture Organisation of the United Nations |
| GAM | Generalised additive model |
| Ha | Hectare |
| IPCC | Intergovernmental panel on climate change |
| IRECI | Inverted Red-Edge Chlorophyll Index |
| LAI | Leaf area index |
| Mg | Megagram (equal to 1000 kilograms) |
| NDVI | Normalised difference vegetation index |
| NRMSE | Normalised root-mean-square error |
| $N_2O$ | Nitrous oxide |
| GNDVI | Green normalised difference vegetation index |
| NIR | Near-infrared |
| LiDAR | Light detecting and ranging |
| MCARI | Modified Chlorophyll Absorption in Reflectance Index |
| R | Red |
| RGB | Red-green-blue |
| RE | Red-edge |
| RMSE | Root-mean-square error |
| SR | Simple ratio |
| SWIR | Shortwave infrared |
| UAV | Unmanned aerial vehicle |
| VI | Vegetation index |

# 1. INTRODUCTION

*Agave sisalana* (sisal), native to Central America, is a perennial succulent, widely introduced to the tropics and subtropics as a crop plant (Singh, 2013). The main product from the sisal crop is a hard fibre extracted from its leaves and used for ropes, clothes, carpets, and as a reinforcing composite i.e. in automotive industry. Global production of sisal peaked over 600 000 Mg in the 1960s, after which competition with synthetics fibres led to sharp decline in its production. In the 2000s, the demand has seen a new growth, due to revived interest in natural, biodegradable fibres (FAO, 2020). According to the Food and Agriculture organisation of the United Nations (FAO) the annual production of sisal in 1998–2018 averaged 320 000 Mg. This places it 6$^{th}$ among natural fibres in terms of production, accounting for about 2% of global plant fibre production and 70% of the world's hard fibres. The total area under sisal cultivation in 1998–2018 was 359 357 ha on average.

Research literature has addressed sisal's physical properties and use as a composite (Sahu & Gupta, 2017), as well as its chemical composition and use in chemical and pharmaceutical industries (Debnath *et al*., 2010; Santos *et al*., 2015). Recently, there has also been a growing research interest for sisal and other *Agave* plants as a biofuel feedstock, reflecting the increasing need for sustainable energy sources (Davis *et al*., 2011; Escamilla-Treviño, 2012; Terrapon-Pfaff *et al*., 2012; Pérez-Pimienta *et al*., 2017; Niechayev *et al.*, 2019). Currently in the sisal industry there is a largely untapped potential to channel the waste-products of the fibre production towards biofuel production, as the fibre comprise only a small part of sisal's biomass and the residue is traditionally considered waste (Singh, 2013). *Agave*'s drought tolerance and high biomass yield, in semiarid and arid environments, could also create opportunities to grow them for biofuel production in marginal lands, labelled unsuitable for agriculture and thus without competition from the food production (Von Cruz & Dierig, 2015).

Currently, there exists no research on quantifying the biomass of sisal. In agriculture, biomass is an important parameter for monitoring the crop status, growth and yield (Serrano *et al*., 2000; Lemus and Lal, 2005; Ahamed *et al.*, 2011). Furthermore, quantifying biomass can be used as a means to estimate carbon sequestration (Zan *et al*., 2001). At field level, allometric equations, which can be used to predict plant's mass from its dimensions, such as the height and diameter, are a practical non-destructive way to assess biomass (Chave *et al*., 2014). Precise predictions, however, require equations that are specific to species or plant functional type (Paul *et al*., 2016). On larger scales, assessing biomass with field measurements is resource-intensive and time consuming. Remote sensing, i.e. measuring objects from a distance with satellites,

provides a solution for repeating and upscaling the biomass estimations in a more efficient way (Ahamed *et al.*, 2011).

Despite the rapid development of remote sensing technology and its growing use in crop biomass mapping (Chao *et al*., 2019), it has not yet been tested for assessing the biomass of sisal, or any other plant in *Agave*-genus. Using remote sensing to assess crop biomass is, however, a relatively quick and cost-effective method and advantageous for two reasons. Firstly, it can be used for yield predictions and for large-scale monitoring of the crops in almost real-time, providing valuable information for resource planning and yield optimization (Ahamed *et al*., 2011; Battude *et al*., 2016; Chao *et al*., 2019; Serrano *et al*., 2000). Secondly, remote sensing crop biomass can be used to derive information of carbon cycle and crop's role in climate change mitigation (Lemus & Lal, 2005; Sartori et al., 2006).

Multispectral satellites, such as the European Space Agency's (ESA) Sentinel-2 have shown their potential for large scale biomass estimation in different environmental contexts (Castillo *et al*., 2017; Sibanda *et al*., 2015) and have also been used to map crop biomass (Battude *et al*., 2016). A common way in multispectral biomass modelling is to calculate spectral transformations, called vegetation indices (VIs), which can be used as indicators of plant biophysical characteristics, such as biomass (Serrano *et al*., 2000). Because of the unique spectral and structural characteristics of different plant species, vegetation indices' sensitivity is often plant specific (Marshall & Thenkabail, 2015). The success of crop biomass models can also vary between species and the environmental context (Prabhakara *et al*., 2015). Species-specific research is thus required, to evaluate the feasibility of such modelling approach for species that have not yet been studied.

In this thesis, the objective was to assess the biomass of sisal. First, by using field-measurements, and then by up-scaling the assessment with remote sensing methods. More specifically there were three objectives: (1) to develop an allometric biomass equation for sisal leaves, (2) assess the utility of medium-resolution multispectral satellite imagery in estimating sisal leaf biomass, and (3) to model sisal leaf biomass at plantation level. These objectives were translated into the following research questions:

1) What are the allometric relationships for predicting the biomass of *Agave sisalana* leaves?
2) What is the relationship between multispectral vegetation indices and the leaf biomass?
3) What is the spatial distribution of the leaf biomass in the study area?

# 2. BACKGROUND

## 2.1 Plant biomass and crops

Biomass is plant material formed when plants absorb the sun's energy through photosynthesis, to convert carbon dioxide ($CO_2$) and water into nutrients (Field & Raupach, 2004). Aboveground biomass (AGB) refers to all the biomass above the ground, or more specifically to all living biomass above the soil including stem, stump, branches, bark, seeds, and foliage (IPCC, 2006). The carbon content of AGB is 45–50%, depending on the plant species (Ma *et al*. 2017). Recent decades have seen a surge of interest in quantifying AGB stocks of various terrestrial biomes, due to their functioning as a part of the global carbon cycle and therefore their integral role in the climate system and its change (Field *et al*., 2008; Saatchi *et al.*, 2011).

In the global carbon cycle plants belong to terrestrial biosphere, which contains the organic carbon stored in land-living-organisms and in the soil. In the terrestrial cycle plants act both as a carbon sink and source (Field & Raupach, 2004). Plants bring carbon to the terrestrial biosphere through photosynthesis and release it back to the atmosphere through respiration, decomposition and fires. Also human actions, such as land use practices e.g. conversion of natural vegetation to croplands cause carbon to be released (Pellikka *et al*. 2018). In addition to the carbon that is sequestered and released in the form of AGB, plants control also the soil organic carbon (SOC) input by translocation carbon trough their root system (Mathew *et al*., 2017). Overall, the soils of the world contain more carbon than the atmosphere and living plants combined (P. Smith, 2006). On long timescales, such as decades or centuries, the balance between the gains and losses determines how much carbon is stored in the all the biosphere. Consequently, this balance has a significant effect on atmospheric $CO_2$ levels and climate system. On average, the terrestrial biosphere has been a carbon sink, but anthropogenic activities such as land-use changes have reduced the total carbon uptake of the biosphere (IPCC, 2019). This, together with increased atmospheric $CO_2$ due to fossil fuel combustion, are the main drivers of the climate change.

Globally, agricultural lands are one of the main SOC stocks (Kell, 2012). Since biomass is one of the key determinants of SOC sequestration, quantifying crop AGB has been used as a means to estimate carbon sequestration in agricultural lands (Lemus and Lal, 2005, Mathew *et al* 2017). However, crop biomass has also the function of food and bioenergy stock. It is therefore a fundamental parameter in agriculture, used as an indicator of crop growth, status, yield and response to agricultural management practices (Lemus & Lal, 2005; Serrano *et al*., 2000). Consequently, the methods for measuring crop biomass non-destructively have been a continuing research interest (Tucker, 1980; Chao *et al*., 2019). Such research has encompassed

both food crops such as maize and wheat (Serrano *et al*., 2000; Battude *et al.*, 2016), as well as energy crops like perennial grasses grown for biofuel production (Youkhana et al., 2017).

## 2.2 Crop biomass assessment

### 2.2.1 Plant allometry

Allometry, a foundation for plant biomass assessment, is a study of plant growth. More specifically, it is a study of the statistical relationships in the dimensions of the plant that are correlated with changes in the overall size and structure, due to the similarity in individual's ontogenetic development (Niklas, 2004). Such dimensional relationships can be employed for biomass estimation, as it has been shown that the proportions between plant height, diameter and biomass follow statistical rules, which are same for certain plants, growing under similar conditions (Chave *et al*., 2014). Once these relationships are established into allometric equations, they can be used to non-destructively estimate plant biomass in the field. The relationship is usually established by regressing a dependent variable (biomass) against one or several independent variables (i.e. plant diameter and height), but also non-linear approaches have been used (Litton *et al*., 2006).

A commonly used approach in allometric biomass modelling is logarithmic-transformation of variables, followed by linear regression (Mascaro *et al*., 2011). Log-transformation often allows the relationship between variables to be described with simple linear equation, while also normalising the error structure. This approach has been criticized by Packard and Boardman (2008), who argued that analyses should be performed on the original scale, since log-transformation can lead to "biased and misleading estimates", as such models operate in geometric rather than arithmetic space. However, others such as Kerkoff and Enquist (2009) and Mascarov *et al*. (2011) have pointed out, that biological growth is "multiplicative by nature" and therefore log-transformation is fully acceptable, since it accounts for proportional rather than an absolute variation.

A vast number of biomass equations that use height, diameter and stem volume as predictors have been formulated for both tropical and boreal trees (Chave *et al*., 2014; Peichl & Arain, 2007). For herbs and shrubs, the stem and foliage diameter have been used as predictors (B. W. Smith & Brand, 1983). In crop allometry, Youkhana *et al*. (2017) used stem diameter in a simple power model to predict napiergrass, energycane, and sugarcane biomass with good accuracy (coefficient of determination ($R^2$) = 0.96–0.99). For soybean, Reddy *et al* (1998) found a log-log linear dependency between both height and diameter and mass. Currently, no research has

been done on sisal allometry and hence it is unknown how the dimensional variables relate to its mass.

As allometric models are formulated based on certain size range determined by the sample minimum and maximum values, using them to predict values outside of this range can result in biased estimates (Picard *et al*., 2014). The allometric theory also states that the models are applicable for plants growing under similar conditions (Niklas, 2004). However, Paul *et al*. (2016) have tested the generality of various allometric shrub- and tree-models across different ecoregions and found no significant bias in their prediction accuracy.

### 2.2.2 Remote sensing

During the recent decades, remote sensing has emerged as a feasible and widely used method to assess biomass (Ahamed *et al*., 2011; Kumar *et al*., 2015). Remote sensing refers to a process of studying the characteristics of objects or areas from a distance, by measuring the reflected and emitted radiation (Jensen, 2014). The most essential remote sensing methods for biomass estimation include satellite and airborne imagery, light detecting and ranging (LiDAR), synthetic aperture radar (SAR) and recently also unmanned aerial vehicles (UAV) (Ahamed *et al*., 2011). These methods, either individually or combined, can be applied to quantify and monitor biomass, from local to regional scales, each with their strengths and weaknesses.

Multispectral satellites, from high-resolution (3–5m) satellites such as RapidEye and PlanetScope, to medium resolution (10–30 m) satellites like Sentinel-2 and Landsat 8, are some of the most common data sources for large scale remote sensing of crop biomass (Ahamed *et al*., 2011). Their strengths include global coverage, good spatial resolution and quick revisit time (1-5 days), which facilitates biomass estimations over large areas with good temporal resolution and precision (Battude *et al*., 2016). Being open access products, Sentinel-2 and Landsat 8 are also accessible even with limited budget. Multispectral satellites have been successfully used in large scale biomass modelling of forests (Castillo *et al*., 2017), grasslands (Sibanda *et al.*, 2015) and crops such as maize (Battude *et al*., 2016), but have not yet been tested for modelling the biomass of *Agave*-crops.

Despite the quick revisit time, the temporal resolution of multispectral satellites is, in reality, limited by cloud coverage, which can make the availability of data unpredictable (Asner, 2001). Another challenge is that the transferability of the multispectral biomass models across regions can be challenging, due to varying external factors across different environments (Foody *et al*., 2003). Because multispectral biomass estimations also require a large number of in situ sample

data, to calibrate, validate and assess the accuracy of the models, building site specific models can be time-consuming and expensive, even if the image data were free (Lu, 2006). Nonetheless, because of simplicity, availability and interpretability, multispectral satellites have become a standard data source for biomass estimation (Chao *et al*., 2019). Especially the new generation satellites such as Sentinel-2, equipped with a sensor designed to be specifically sensitive to vegetation characteristics, have proven their suitability (Sibanda *et al*., 2015; Battude *et al.*, 2016). In addition to their applicability for simple multispectral models, multispectral data can be used in synergetic multi-source models, that combine different data types (Chang & Shoshany, 2016), or multiple spatial scales (Gnyp *et al*., 2014; Riihimäki et. al. 2019). Furthermore, multispectral data can be fed into crop models, which are used to predict the growth and production (Claverie *et al*., 2012).

### 2.2.3 Principles of multispectral satellite remote sensing

Multispectral satellites orbiting the Earth are equipped with sensors that observe and record information from the Earth's surface, without direct contact to it. This information, whether it is the condition, structure, or state of an area or object, is transmitted to the sensor in the form of electromagnetic radiation (Eamu *et al*., 2016). The basis of an accurate interpretation and effective application of multispectral data is in understanding the physical principles of how the electromagnetic radiation interacts with the Earth's surface. It is this interaction that provides us information about the surface, when we interpret the signal the sensor has detected.

The electromagnetic radiation can be conceptualized as continuous waves (or frequencies) which transmit energy from one location to another (Jensen, 2014). The distance between two successive peaks of a wave is called a wavelength. The whole range of wavelengths (or frequencies) constitutes the electromagnetic spectrum, ranging from gamma rays to radio waves (Fig. 1). Only a small portion of this spectrum is visible, or detectable with human eye. The basis of remote sensing is that the electromagnetic signal, received by the sensor, will be uniquely different across the wavelength range for different types of surfaces on the Earth.

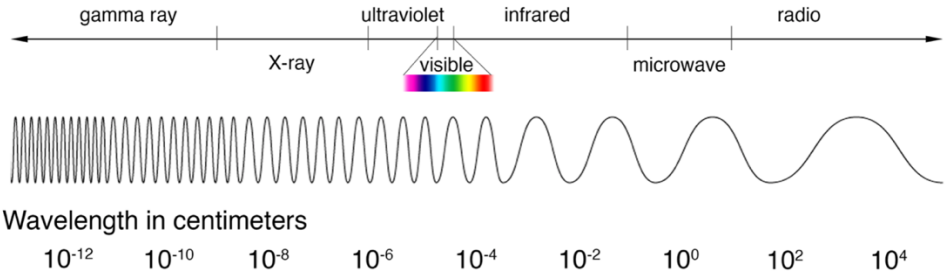


*Fig. 1. Electromagnetic spectrum (NASA, 2013).*

The source of the electromagnetic radiation that a passive remote sensing sensor detects is the Sun (Eamus *et al*., 2016). When the Sun's energy is incidence to the surface of the earth, three processes occur: the energy can either be reflected, transmitted or absorbed. The proportions of these processes are a function of the unique characteristics of the surface and differ with wavelength. These variations provide us information about the surface, by giving objects or areas their unique spectral signatures. The spectral signatures are produced by the optical properties of the biophysical and biogeochemical constituents, moisture condition, and the size, shape and geometry of an object. With an accurate interpretation of these signatures, we can derive information about physical, chemical, and biologic state of the objects and surfaces. It has been shown, for example, that by analysing the spectral properties of plants, we can accurately assess their nitrogen and chlorophyll content (Schlemmera *et al*., 2013), relative water content (Raymond *et al*., 1987), species richness (Gould, 2000), as well as biomass (Serrano *et al*., 2000; Li *et al.*, 2016; Wang *et al.*, 2016).

The Sun, the source of the electromagnetic radiation, emits energy across most of the electromagnetic spectrum, at wavelengths of gamma rays (~0.0001 nm) to radio waves (~10 m) (Jensen, 2014). Multispectral satellites are equipped with an optical sensor, which means they detect information at wavelengths ranging from the visible to the shortwave infrared region of the electromagnetic spectrum (400–2500 nm). Depending on the sensor, this wavelength range is furthermore split into a varying number of spectral bands (also known as channels), which detect the signal at specific wavelength ranges. These bands usually include at least blue (central wavelength near 490 nm), green (560 nm), red (665 nm) and near-infrared (NIR) (830 nm) bands, and varying number of shortwave infrared (SWIR) bands (1500–2000 nm). By exploring the differences of object's signal on different wavelength bands, we can acquire information about the characteristics of vegetation. An effective method for doing this is the calculation of spectral transformations, known as vegetation indices (Silleos *et al*., 2006)

### 2.3.4 Plant spectra and vegetation indices

The absorption of a green vegetation is controlled by its chemical, moisture and physical properties (Eamus *et al*., 2016). Leaf pigments (chlorophylls, carotenoids, xanthophylls and anthocyanins) absorb electromagnetic radiance in the visible wavelengths (400–700 nm) for photosynthetic purposes, which results in low reflectance at those wavelengths (Fig. 2). Relatively less absorption occurs in the green wavelength (500–600 nm), than in blue (450–500 nm) and red (600–750nm), creating a reflectance-peak in the green spectral region, which is the

reason why plants are perceived as green by the human eye. In healthy green leaf the reflectance in the NIR region (700–1200 nm) typically increases drastically, because plants do not use this energy and have therefore adapted to reflect and transmit it, to avoid warming that would harm the plant. This process is also referred to as NIR-scattering. The water content of the leaf likewise affects the reflectance properties since water is a strong absorber of energy. This is evident on all wavelengths, but especially in SWIR region (1300–2500 nm), where reflectance decreases the most with increasing moisture. Because of reflectance responses within visible, near-infrared and middle-infrared regions and their relation to vegetation's biophysical state, data from these regions is used in remote sensing to calculate vegetation indices.

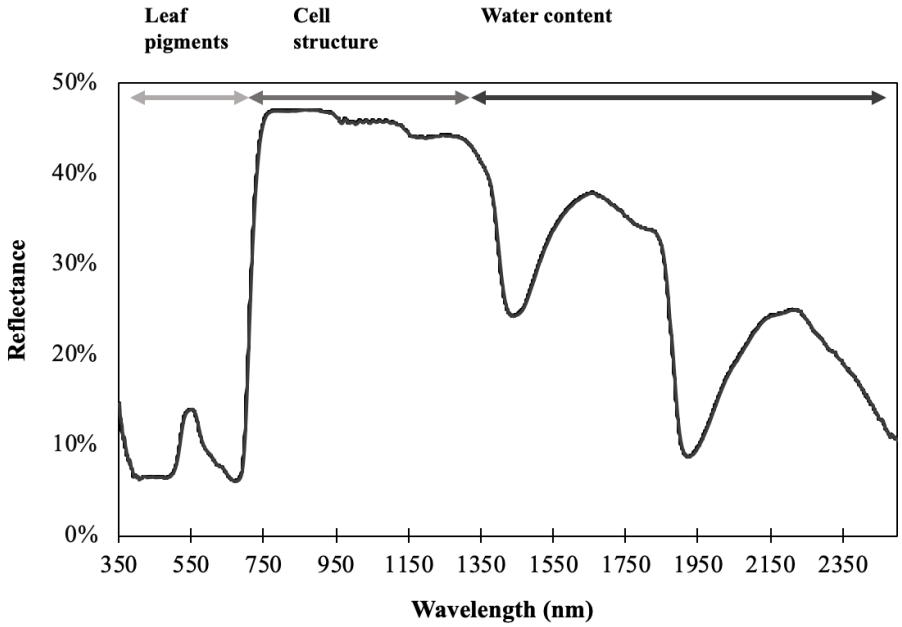


*Fig. 2. Leaf spectra of Arrhenatherum (Helsen et al., 2020) and the main determinants of vegetation spectral (Eamus et al., 2016). Spectral data retrieved from an open dataset by Helsen et al. 2020.*

Vegetation indices (VIs or more specifically broadband VIs in the case of multispectral sensors) are dimensionless, radiometric values that indicate crop biophysical characteristics such as abundance and activity of green vegetation, chlorophyll content, absorbed photosynthetically active radiation and biomass (Jensen, 2014). When plants grow and photosynthetic activity increases, the reflectance in the red wavelength region decreases, while the NIR reflectance increases (Eamus *et al*., 2016). Therefore, signal from these two regions has been widely used to formulate VIs, such the Simple Ratio (SR) (Birth & McVey, 1968), which is simply the ratio of red reflectance ($\rho_{red}$) to near-infrared reflectance ($\rho_{nir}$):

$$SR = \frac{\rho_{red}}{\rho_{nir}}$$

Another widely used VI is a nonlinear transformation of SR called normalized difference vegetation index (NDVI) (Rouse *et al*., 1974):

$$\mathrm{NDVI} = \frac{\rho_{\mathrm{nir}} - \rho_{\mathrm{red}}}{\rho_{\mathrm{nir}} + \rho_{\mathrm{red}}}$$

NDVI is functionally equivalent to SR and they are both known to be sensitive to biomass variations of e.g. wheat (Serrano *et al*., 2000), corn, soybean (Kross *et al*., 2015) and barley (Bendig *et al*., 2015). NDVI tends to saturate at high biomass values (Goswami *et al.*, 2015), but this can be overcome if the bands used are narrow enough (Mutanga & Skidmore, 2004). In addition to NDVI and SR, a vast number of VIs have been used for biomass estimation (Silleos *et al.*, 2006; Panda *et al*., , 2010; Tilly *et al*., 2015), but most of them are essentially transformations of red and NIR reflectance.

VIs used for the assessment of leaf chlorophyll content and leaf area index (LAI, one-sided leaf area per unit ground) have utilized also the reflectance from the green spectral region (Gitelson *et al*., 2011). This can be applicable for biomass estimation as well, as these plant parameters are highly related (Filella & Peñuelas, 1994). In addition to ratio and normalized ratio transformations, also the subtraction of reciprocal reflectance's has proven to be useful for chlorophyll estimation. For instance, anthocyanin reflectance index:

$$\mathrm{ARI} = \frac{1}{\rho_{\mathrm{green}}} - \frac{1}{\rho_{\mathrm{red\text{-}edge}}}$$

which Merzlyak *et al*. (2003) found to be proportional to anthocyanin concentration.

An important wavelength region for VIs is also the red-edge, a rapid increase between red and near-infrared wavelengths (700-800 nm) (Jensen, 2014). Horler *et al*. (1983) have shown that the red-edge position, the point of the maximum slope (or the first derivate of reflectance) in the red-NIR region, is mainly controlled by chlorophyll concentration, with additional effects from species, developmental stage, leaf layering and leaf water content. Thus, it can be used as an indicator of plant status (Filella & Peñuelas, 1994). Because the red-edge position varies between plants and growth stages, the new generation multispectral satellites such as Sentinel-2 have narrow red-edge bands between the red and NIR wavelengths, positioned at slightly different central wavelengths. Recent studies have shown that Sentinel-2 red-edge bands have the potential to slightly increase the precision biomass estimations, compared to other

multispectral satellites and other Sentinel-2 bands (Sibanda *et al*., 2015; Forkuore *et al*., 2018; Pandit *et al*., 2018). However, there are also examples where the red-edge indices did not result in significant improvement (Kross *et al*., 2015).

Because of the unique spectral features of different crops, as well as varying external factors, models that use multispectral VIs to estimate crop biomass have shown performances ranging from low to high. For example, Prabhakara *et al*. (2015) achieved $R^2$ of 0.86 for different winter crops using hand-held spectrometer, red and NIR based VIs and regression models. For corn and soybean, Kross *et al*. (2015) used RapidEye NDVI and red-edge based VIs in regression models, which resulted in $R^2$ of 0.86–0.88. For wheat biomass, Wang *et al*. (2016) used random forest regression and multiple HJ-1 (Huan Jing-1: Environmental Protection & Disaster Monitoring Constellation) based VIs and achieved $R^2$ of 0.79 between the observed and predicted biomass. The performance of VIs in assessing sisal biomass is not known, since multispectral data has not been applied for that purpose.

# 3. STUDY AREA

## 3.1 Teita Sisal Estate

The study area – Teita Sisal Estate (3°30′ S, 38°24′ E) – is located in the town of Mwatate, in Taita-Taveta County in the Coast Province of Kenya, and right next to the Taita Hills (Fig. 3). Taita Hills are the northernmost part of the Eastern Arc Mountains, a chain of ancient mountains across the Eastern parts of Tanzania and Kenya (Pellikka *et al*., 2013; Platts *et al*., 2011). The semi-arid climate in the area is determined by the tropical location, variations in the elevation and proximity to the Indian Ocean (Fig. 4). Rainy seasons take place in March-June and October-December and orographic rains cause variability in precipitation throughout the year. According to the Teita Sisal Estate's own rainfall record, the mean annual rainfall in 1960–2018 was 610mm.

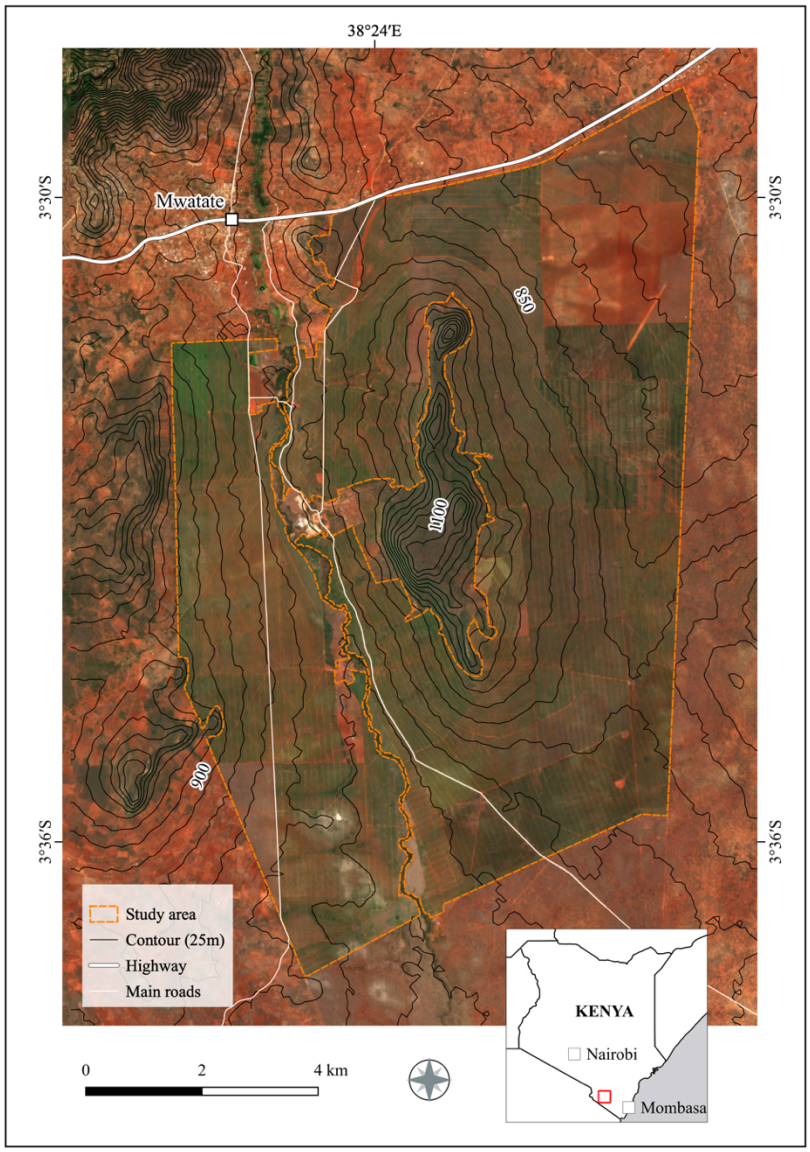


*Fig. 3. Location of the study area (Teita Sisal Estate). Sentinel-2 RGB image from 28.9.2019 as a base map.*

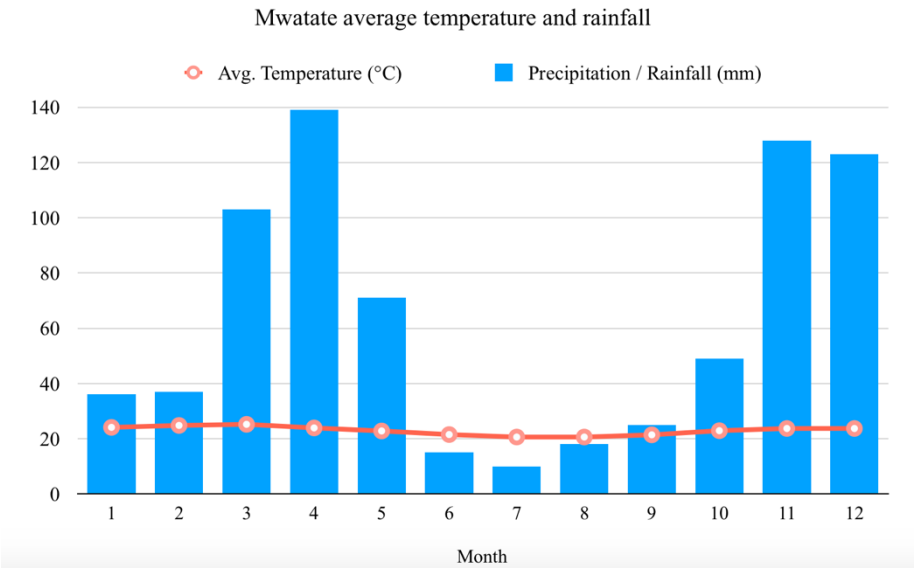


Fig. 4. Average temperature and rainfall by month in Mwatate in 1982 – 2012. Source: https://en.climate-data.org.

The Estate is an important employer in the area, employing up to 2000 people on average (E. Mrombo, personal communication, 19 May, 2020). It is also one of the largest sisal plantations in the world, with a cultivable area of 8851ha. The monthly fibre production is 800 Mg on average, which means it accounts for over one third of the total sisal production in Kenya, the third largest sisal producing country in the world. Total sisal fibre production in Kenya in 1998–2019 averaged 22 975 Mg a year (FAO, 2020).

## 3.2 The cultivation of sisal

Most of the Estate's 8851ha of cultivable area is in active use. Elevation of the fields ranges 700–950m above sea level and the slope is 0–7°. Mainly three different sisal varieties are cultivated at the plantation: *Agave sisalana*, *Agave hildana* and *Agave hybrid 11648*, of which the last one occupies most of the area (E. Mrombo, personal communication, 19 June, 2019). The crop is grown from small saplings, with a plant population of 4995 per hectare. The management practices differ between the field blocks. Generally, herbicides are applied for young plants for weed control, while the weed control for the older plants is minimal and some of the fields have not received herbicides at all. Also grazing (mainly cows) is practiced in the fields. Varying management practices mean the field blocks are heterogeneous in terms of the understory vegetation (Fig. 5). Younger fields have no or just little weeds, while the older fields have a varying cover of weeds and shrubs. Also the fertilization practices are varying. Some fields have received fertilizers, such as NPK 171717 (nitrogen, phosphorous, potassium) and sisal waste, produced during the fibre extraction, while other fields have not received fertilizers.

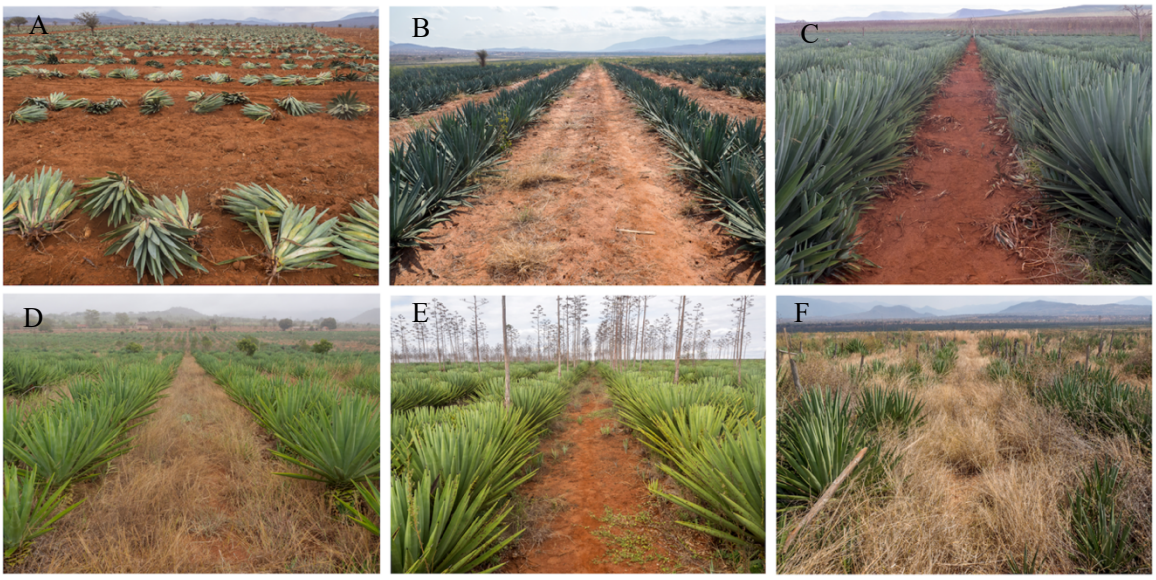


*Fig. 5. (A) Sisal saplings ready to be planted. (B) 1-year-old plants with little understory. (C) 3-year-old plants with no understory. (D) 7-old-plants and dry weeds. (E) 13-year-old plants, with flower stalks and little understory. (F) Old field, where the weeds have taken over.*

Sisal plant forms a rosette of leaves around its stem. The harvesting starts at the age of 2.5–3 years when the oldest leaves, lowermost in the rosette, are cut manually. After harvesting, the leaves are taken to the processing site where the fibre is extracted, dried in the sunlight and prepared for shipping (Fig. 6). Harvesting continues regularly up to 10–15 years, after which the plant grows a flower stalk, which can also be harvested, and used as a construction material. What is left after all the leaves and the stalk have been harvested, is a leftover stump, which consists of a stem, bits of leaf bases and the base of a flower stalk (Fig. 7). It is also referred to as sisal ball and used as a manure by burning and ploughing it into the soil (Terrapon-Pfaff *et al*., 2012).


*Fig. 6. (A) Mature field, where the harvesting has started. (B) Leaves are transported to the processing site for fiber extraction. (C) The fiber is dried in the sunlight. (B) Processed fiber ready to be shipped.*

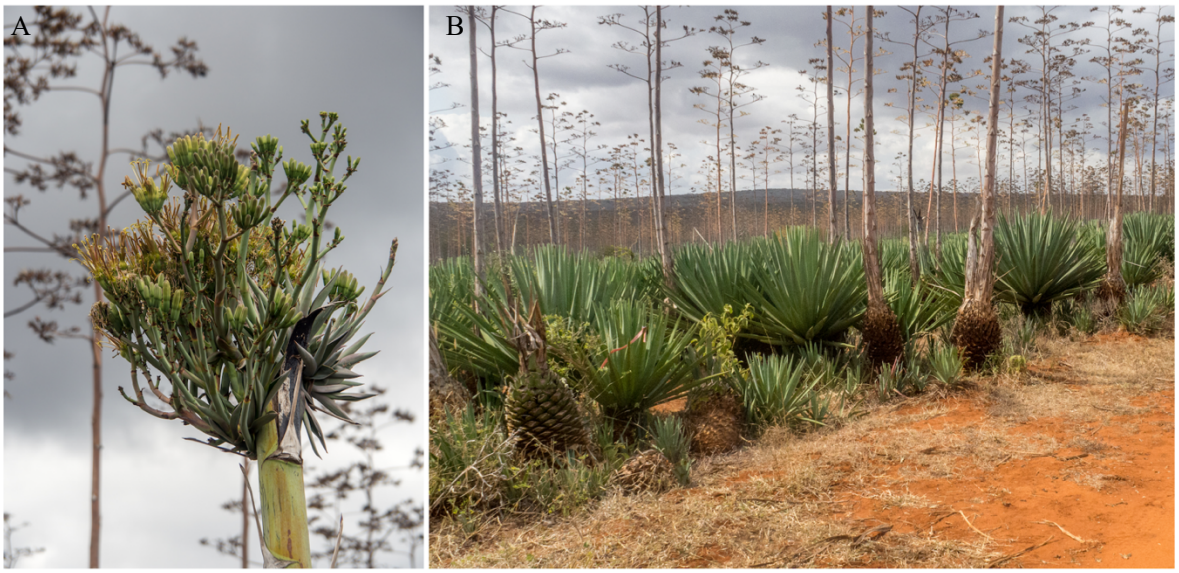


*Fig. 7. (A) Sisal flower. (B) Mature sisal field with flower stalks and sisal balls by the road.*

# 4. DATA & METHODS

Data and methodology are summarized as a workflow in Fig. 8, which also presents the resulting outputs. Subsequent chapters describe the workflow in a more detail.

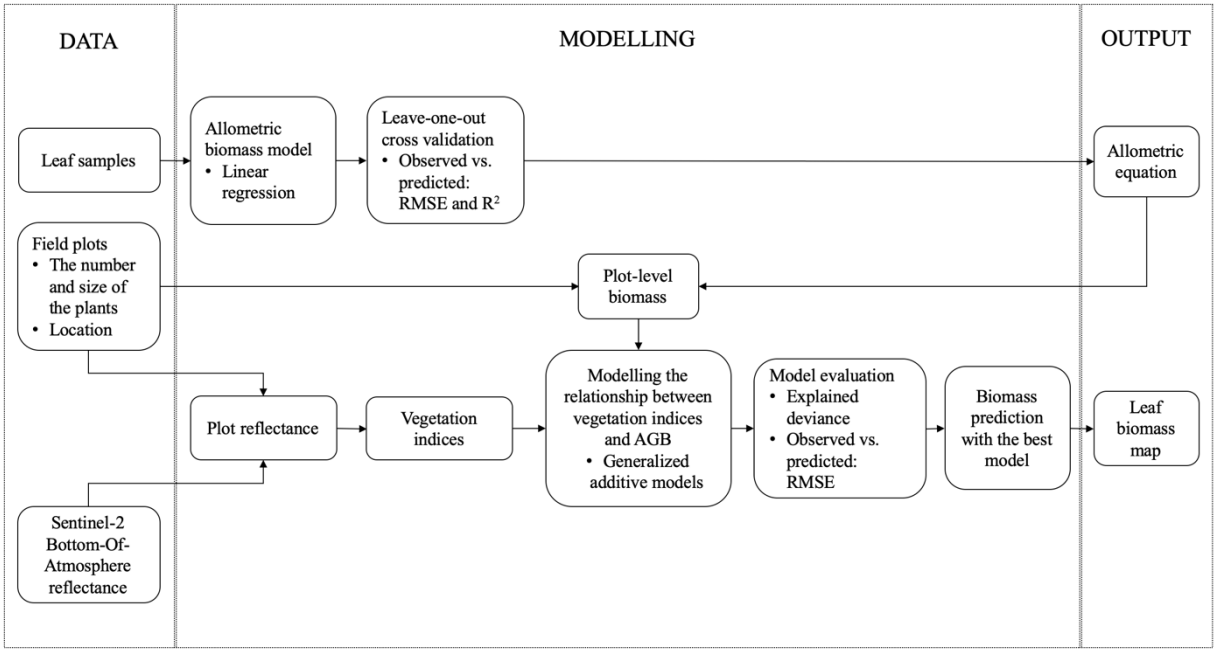


*Fig. 8. Research workflow, which shows the data, main methodology and the resulting outputs.*

## 4.1 Biomass assessment at leaf and plot level

### 4.1.1 Leaf sampling

For the development of the allometric model a sample of sisal leaves (n = 38) was harvested in August 2019 (Fig. 9). All the leaves were harvested from individual plants and the sample included two sisal varieties: *Agave hybrid* and *Agave hildana* (Fig. 10). Sampling covered only the leaves, since they are the main stock of sisal's biomass and the part of the plant that is regularly harvested during its 10–14-year life cycle. Picard *et al*. (2012) suggest stratified sampling as a sound approach to explore the variability in a study area and to increase the precision of the modelling. Therefore sites (i.e. land parcels) for the sampling were subjectively chosen based on NDVI-values, as well as prior knowledge of the planting time, to cover the whole range of leave sizes. Plants in the proximity of roads were avoided, but otherwise the plants for the sampling were chosen randomly at the sites. Plant height, number of overlapping leaves in the rosette, and the position of the leaf within the rosette were recorded. To standardise the measurements, all the leaves were cut along the narrowest width at the leaf base.


*Fig. 9. Measuring leaves for allometric modelling. (A) The whole leaf sample. (B) Sub-sampling a leaf for drying. (C) Sub-sample weighing. (D) Subsamples placed in the oven for drying.*

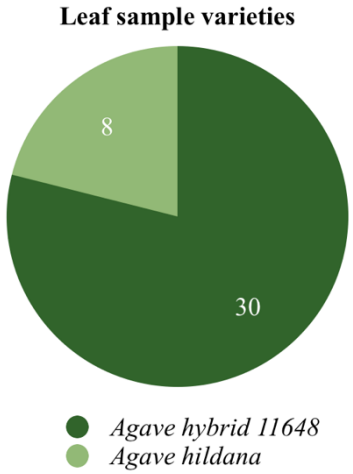


*Fig. 10. Distribution of the two sisal varieties in the sample.*

After harvesting, the leaves were taken to a laboratory at Taita Taveta University to be measured and dried. A few chosen leaves were weighed immediately after cutting and at the laboratory, to ensure that the leaves had not lost any weight by evaporation while being transported. At the laboratory the leaves were first weighed for their fresh weight with a table-scale (d = 0.01g) and measured for their length and maximum diameter. Before drying, the leaves were subsampled to fit all the samples into the oven. To account for possible density differences along the leaf axis, the subsampling was done by dividing each leave into four parts of equal length and by taking 5 cm samples from the middle of each part (Fig. 9B). The samples were then weighed, placed on dishes and put into the oven to 70Cº (Fig. 9C-D). Chosen samples were weighed every day until a constant weight was reached after 72 hours. After this, all the samples were measured for their dry weight. The dry weight to wet weight ratio was then calculated for all the samples and used to calculate the dry weights of the whole leaves.

### 4.1.2 Allometric modelling

The relationship between the leaf biomass and dimensions was formalized with linear regression, which is suggested as an appropriate method when the aim of the allometric modelling is prediction (Warton *et al*., 2006). Two linear regression models were fitted in R-Studio version 1.2.5019 (R Core Team, 2020) using the least squares method. Dry biomass was used as a response variable and leaf maximum diameter (D), plant height (H) and leaf length (L) were combined to $D^2H$ and $D^2L$, to represent volume approximations and used as single explanatory variables. Natural log-transformation was used for both response and explanatory variables. Such transformation is convenient, as it can render the relationship between variables linear, and tune the error structure of the model, to meet the assumptions of linear regression (Picard *et al*., 2012). This is often applicable for allometric models, which tend to have multiplicative error structure. Consequently, the linear models were fitted as

$$\log(Y) = \beta_o + \beta_1 \log(X) + \varepsilon$$

where $X$ is the explanatory variable, $Y$ the response variable, $\beta_o$ the y-intercept the line, $\beta_1$ its slope and $\varepsilon$ the error term.

The assumptions of linear regression were tested visually with diagnostic plots and more formally with Breusch-Pagan and Shapiro-Wilk tests for normality and homoskedasticity (Breusch & Pagan, 1979; Shapiro & Wilk, 1965). Furthermore, the models were cross-validated using the leave-one-out method (Ruppert, 2004) and by calculating the root mean square error

(RMSE) and $R^2$ between the predicted and observed values. In the leave-one-out cross validation one observation at the time is left out and used as a validation set, while the model is trained with all the other observations. The process is repeated until all the observations have been used for validation, which results in a prediction size equal to the sample size. RMSE is calculated as

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{n}}$$

where y is the observed value, $\hat{y}$ is the predicted value, and n is the number of observations. $R^2$ is calculated as

$$R^2 = \sqrt{\frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2}}$$

where $\hat{y}$ is the predicted value of y and $\bar{y}$ is the mean value of y. Before the RMSE and $R^2$ were calculated the predicted values were transformed back from a logarithmic scale to the original scale with bias correction recommended by Baskerville (1974). The correction factor (CF) is calculated as

$$\text{CF} = \left(\frac{SE}{2}\right)^2$$

where $SE$ is the standard error of the regression. All the figures for this and the subsequent models were produced using ggplot2-package (Wickham, 2016) in R-Studio.

### 4.1.3 Field plots

Field plots (n = 58) were established at Teita Sisal Estate between August 22 and 29, 2019 (Fig. 11). Locations of the plots were chosen subjectively based on NDVI values calculated from a Sentinel-2 image (acquisition date 16 April 2019) and prior knowledge of the planting age, to cover the range of plant sizes and spectral properties in the study area. Furthermore, seven of the plots were positioned next to a gas-chamber measurements sites (at these sites $CO_2$, $N_2O$ and $CH_4$ fluxes were measured over 12-month period for a University of Helsinki research, to be published). At the plantation sisal is planted in double-rows, with a 3.75m spacing between

double rows, 0.7m between single rows and 0.9 m between the plants (E. Mrombo, personal communication, 19 June, 2019). The 20m$^2$ square plots were established with two sides parallel to the rows, so that four double-rows were inside the plot (Fig. 11). Plot locations were recorded with a GNSS receiver (Trimble GeoXH) by measuring the centre and the four corners of the plot. On average, the location of the centre was logged 326 times. A differential correction was applied to the measured locations, in relation to a GNSS base station recording. The location of base station was determined using Trimble RTX post-processing service. Plot polygons (20m$^2$) were digitized afterwards in QGIS 3.4.5 (QGIS Development Team, 2020) using a square-drawing tool. Centre location was used as the exact centre of the polygon and the orientation was guided by the corner locations.

Proximity to roads was avoided and homogeneity preferred, when choosing the plot locations. The number of plants in the two midmost double-rows was counted and multiplied by two, in order to estimate the total number of the plants in the plot. Plant height, the number of leaves, leaf length and maximum diameter were measured from one subjectively determined median sized plant in the two midmost rows. Mean values of these two plants were calculated, in order to constitute representative plot-specific plant metrics. If a considerable variance in plant sizes was observed, then plants were divided into two size-class categories which were measured separately. Some sites had shoots growing on the ground and they were included only if they were higher than 50cm. Furthermore, presence of the understory vegetation and flower stalks, as well as the harvesting status were noted. Information on plant age and variety was received from plantation's books (E. Mrombo, personal communication, 19 May, 2020).

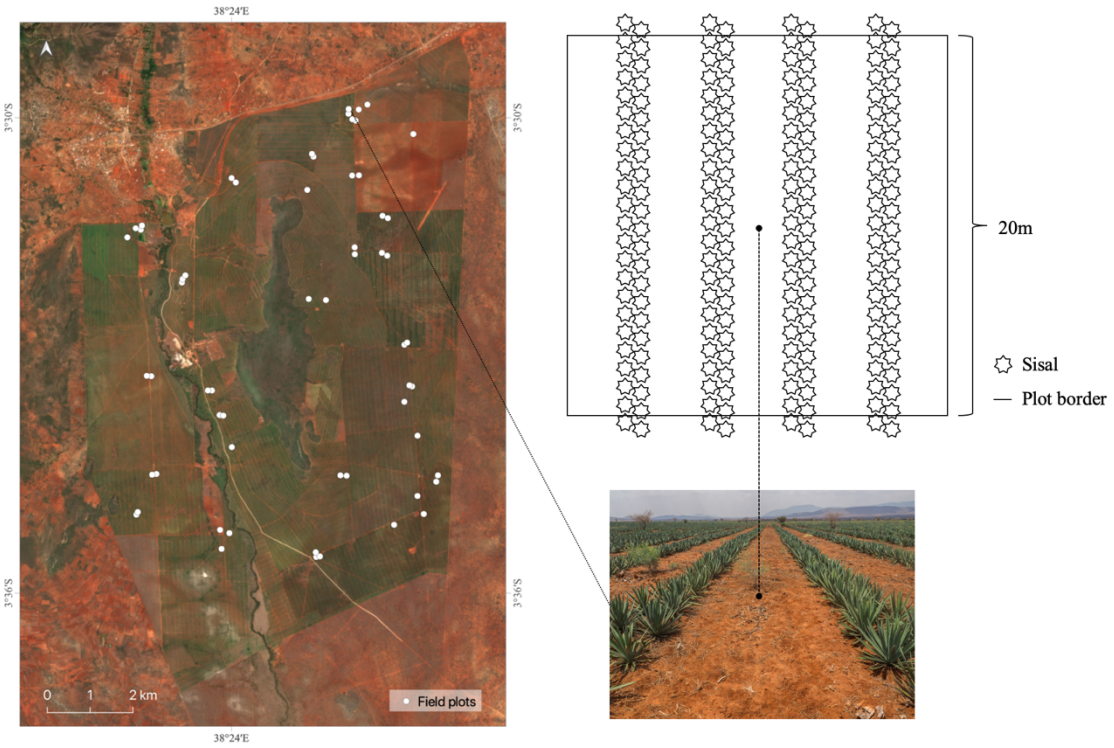


*Fig. 11. Field plot locations on top of Sentinel-2 RGB-image and the plot design.*

Plot level biomass was predicted in R-Studio using the new allometric model (see chapter 5.1.1). First, the plot specific plant metrics (plant height and leaf maximum diameter) were used to predict the biomass of a representative leaf, which was then multiplied by the number of leaves in the plant, and then by the number of the plants, to estimate the total leaf biomass in each plot.

## 4.2 Biomass modelling and mapping using Sentinel-2 data

### 4.2.1 Sentinel-2 satellite image data

Sentinel-2 is a satellite mission developed and operated by the European Space Agency (ESA). It is a constellation of two satellites, Sentinel-2A and Sentinel-2B with a 5-day revisit time. The satellites were launched in 2015 and 2017. The MSI (Multi-Spectral Instrument) sensor aboard Sentinel-2 satellites has 13 spectral bands with spatial a resolution of 10-60m (Table 1).

Sentinel 2 image with acquisition date 9 September 2019 was downloaded from Sentinel open access hub as level 2-A product (https://scihub.copernicus.eu/). Of the cloud-free Sentinel-2 images from the study area, this was the one with the nearest date to the field data collection. Level 2-A products are analysis-ready bottom of atmosphere (BOA) reflectance images,

corrected with the SEN2COR procedure (Sentinel-2 User Handbook, 2019). The bands used were all the bands with 20m spatial resolution (B2, B3, B4, B5, B6, B7, B8A, B11, B12). In the subsequent chapters, they will be referred to as B, G, R, RE1, RE2, RE3, NIR, SWIR1 and SWIR2. Sentinel-2 has two NIR bands and the one used here was the narrow near infrared band (8A).

*Table 1. Sentinel-2 spectral bands.*

| Band | Spectral Band | Central wavelength (nm) | Band width (nm) | Spatial resolution |
|---|---|---|---|---|
| B1 | Coastal aerosol | 443 | 20 | 60 |
| B2 | Blue, B | 490 | 65 | 10 |
| B3 | Green, G | 560 | 35 | 10 |
| B4 | Red, R | 665 | 30 | 10 |
| B5 | Red-edge, RE1 | 705 | 15 | 20 |
| B6 | Red-edge, RE2 | 740 | 15 | 20 |
| B7 | Red-edge, RE3 | 783 | 20 | 20 |
| B8 | Near infrared | 842 | 115 | 10 |
| B8A | Near infrared narrow, NIR | 865 | 20 | 20 |
| B9 | Water vapor | 945 | 20 | 60 |
| B10 | Shortwave infrared | 1380 | 30 | 60 |
| B11 | Shortwave infrared, SWIR1 | 1910 | 90 | 20 |
| B12 | Shortwave infrared, SWIR2 | 2190 | 180 | 20 |

Plot specific reflectance values were calculated from the Sentinel-2 image by taking an area weighted average of the pixels that fell under the plot polygon. The pixel weights were based on the area of the pixel inside the plot polygon. The calculation was done with purpose-build script in Python programming language (Python Software Foundation, 2019)

### 4.2.2 Modelling biomass with vegetation indices

VIs were calculated from the plot reflectance values of Sentinel-2 data. All possible two band combinations of three common vegetation index forms were calculated: 1) ratio based spectral index ($\mathrm{RS} = \frac{Band\ A}{Band\ B}$), 2) normalized difference spectral index ($\mathrm{NDSI} = \frac{(Band\ A - Band\ B)}{(Band\ A + Band\ B)}$) and 3) reciprocal difference vegetation index ($\mathrm{RDSI} = \frac{1}{Band\ A} - \frac{1}{Band\ B}$). In addition, a selection of published VIs were tested as a reference (Table 2.). These included NDVI and SR, which are traditional NIR to red ratios, still widely used in vegetation studies (Ahamed *et al*., 2011). OSAVI and EVI are also based on NIR to red ratio, but with additional parameters that are meant to account for atmospheric and background effects. Rest of the reference indices (CCCI,

IRECI, MCARI, S2REP) take advantage of the bands which are positioned in the red-edge spectral region, usually near 700 nm (Gitelson *et al*., 1996).

*Table 2. Vegetation indices used as a reference.*

| Index | Formula | Reference |
|---|---|---|
| CCCI | $\frac{\frac{NIR - RE1}{NIR + RE1}}{\frac{NIR - R}{NIR + R}}$ | (Barnes *et al*., 2000) |
| EVI | $2.5\,\frac{NIR - RED}{(NIR + 6RED - 7.5BLUE) + 1}$ | (Huete *et al*., 1999) |
| IRECI | $\frac{(NIR - R)}{\left(\frac{RE1}{RE2}\right)}$ | (Frampton *et al*., 2013) |
| MCARI | $\left((RE1 - R) - 0.2(RE1 - G)\right)\left(\frac{RE1}{R}\right)$ | (Daughtry *et al*., 2000) |
| NDVI | $\frac{NIR - RED}{NIR + RED}$ | (Rouse *et al*., 1974) |
| OSAVI | $\frac{(NIR - R)}{(NIR + R + 0.16)}$ | (Rondeaux *et al*., 1996) |
| SR | $\frac{NIR}{RED}$ | (Birth & McVey, 1968) |

The relationships between the biomass and VIs were analysed with Generalized Additive Models (GAM) in R-Studio using *mgcv*-package (Wood, 2017). With GAMs, there is no need to identify polynomial terms or predictor transformations to improve model fit (Wood, 2017). They are also flexible in approximating responses and have relaxed assumptions of predictor-response relationship. Thus, they have the potential to achieve better fits than purely parametric models. Without prior knowledge of the relationships between the leaf biomass and numerous VIs (61 in total), this modelling approach seemed plausible. In remote sensing GAMs have been used e.g. to model fractional vegetation cover (Riihimäki *et al*., 2019) and leaf-area-index (Korhonen *et al*., 2017), with multispectral data.

GAM is a semiparametric generalized linear model that fits the response curves as a sum of smoothing functions (Wood, 2017):

$$y = \beta_o + \sum_{j=1}^{p} f_j(\mathrm{X}_j) + \varepsilon$$

where $y$ is the response vector, $\beta_o$ is the model intercept, $f_j(\mathrm{X}_j)$ the smooth function and $\varepsilon$ is the residual. In GAMs, the relationship between the linear predictor and the mean of the dependent variable is provided by a link function. Here, Gaussian-error structure was used with an identity link function. Smoothing function, which sets the upper limit on the degrees of freedom associated with the smoothing, was set to k = 3, which is considered conservative and should avoid overfitting (Wood, 2017). The models were evaluated based on deviance explained ($D^2$), which is calculated as

$$D^2 = \frac{(Null\ deviance - Residual\ deviance)}{Null\ deviance}.$$

The modelling included two steps: 1) GAMs were calculated for all the VIs, and 2) the reference VIs and two VIs with highest $D^2$ from each group (RS, NDSI, RDSI) were selected for further inspection. The performance of these indices was tested with leave-one-out cross-validation and by calculating RMSE between observed and predicted values. Also normalized RMSE (NRMSE %) was calculated as

$$NRMSE = \frac{RMSE}{\bar{y}} * 100$$

where $\bar{y}$ is the mean of the predicted biomass. NRMSE facilitates the comparison between models with different scales.

# 5. RESULTS

## 5.1 Leaf and plot biomass

### 5.1.1 Allometric model for sisal leaf biomass

Summary of the sampled leaves is shown in Table 3. The water content of the leaves varied between 87–76%. Visual exploration showed that the leaf variables had non-linear relation to leaf dry biomass, all with non-constant variances (Fig. 12). When the single variables were combined into $D^2L$ and $D^2H$ and log transformed, their relation to log(biomass) became linear and the variance close to constant (Fig. 13).

*Table 3. Summary of the sampled sisal leaves. DW = dry weight, WW = wet weight.*

| | Wet weight (g) | Dry weight (g) | DW to WW ratio | Length (cm) | Maximum diameter (cm) | Plant height (cm) | Plant age |
|---|---|---|---|---|---|---|---|
| 1st. Quantile | 111.74 | 21.5 | 0.15 | 57.2 | 16.3 | 84.4 | 3 |
| Median | 272.3 | 53.8 | 0.17 | 81.2 | 20.7 | 140.0 | 8 |
| Mean | 330.82 | 56.3 | 0.17 | 79.2 | 19.9 | 139.1 | 8 |
| 3rd. Quantile | 503.9 | 89.5 | 0.19 | 105.8 | 24.5 | 189.3 | 13 |
| Maximum | 839.22 | 159.0 | 0.24 | 132.5 | 28.4 | 250.0 | 14 |

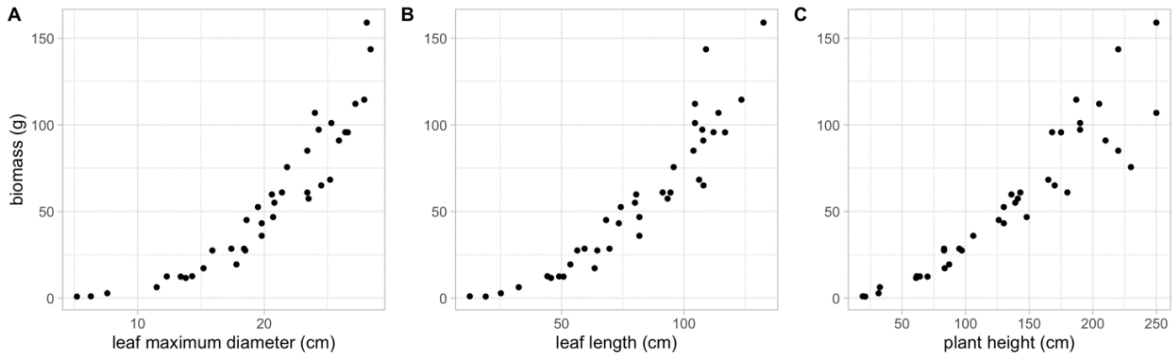


*Fig. 12. The relationship between sisal leaves dry biomass and (A) maximum diameter, (B) length and (C) plant height.*

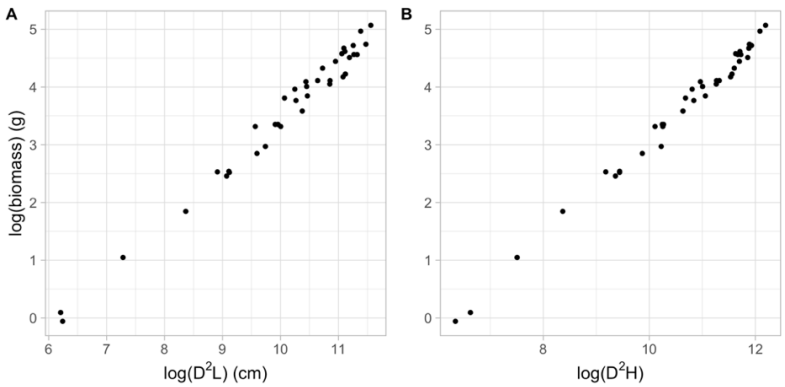


*Fig. 13. Relationship between log(biomass) and (A) log(D²L) and (B) log(D²H) of the leaves.*

Two linear regressions were then fitted to predict log transformed dry weight from (1) log transformed $D^2L$ and (2) log transformed $D^2H$ (Appendix 1). Model 2 (log(Biomass) = log($D^2H$)) had slightly higher adjusted R-squared ($R^2 = 0.99$) and lower residual standard error (RSE = 0.1217) than model 1 (log(Biomass) = log($D^2L$), $R^2 = 0.98$ and RSE = 0.17). Both models were statistically significant ($p < 0.05$). Model diagnostic plots (Fig. 14) showed that the residual structure of the model 1 was skewed and thus violating the assumptions of linear regression. Model 2 had a slight structure with residuals diverging from normal distribution at low and high values, but the assumptions can still be considered to be suitably met. Also, Breusch-Pagan and Shapiro-Wilk test results ($p < 0.05$) indicated that model 2 residuals had close to equal variance and normal distribution.

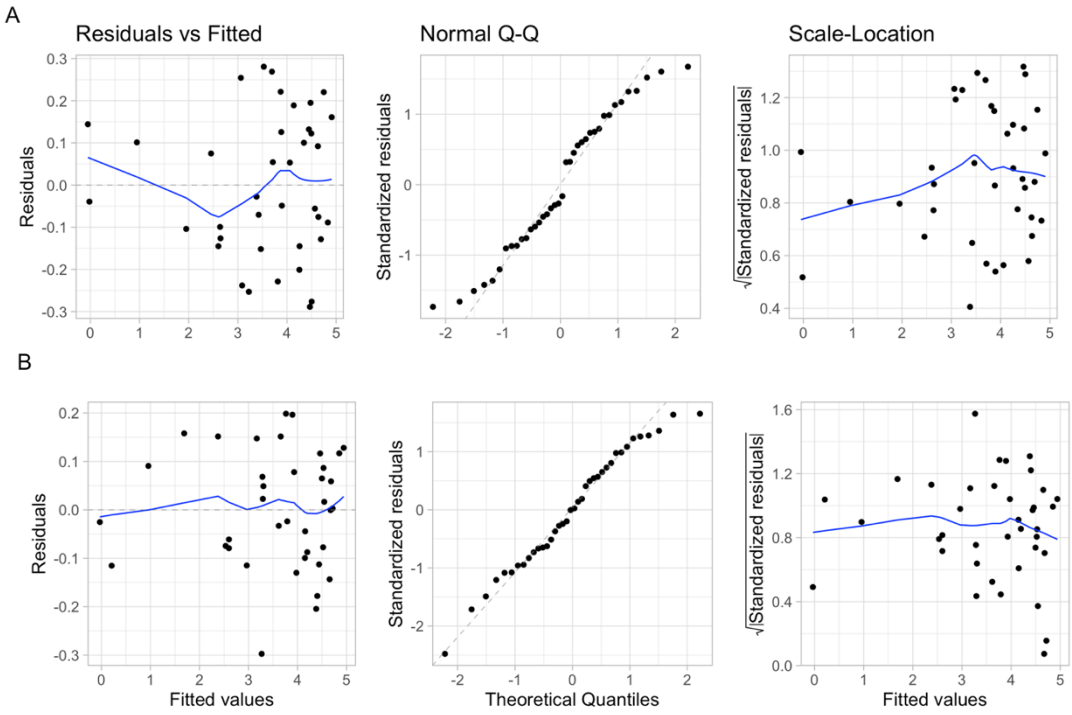


*Fig. 14. Regression diagnostics for (A) log(Biomass) = log(D²L) and (B) log(Biomass) = log(D²H)*

Finally, both models were cross-validated using leave-one-out method, which resulted in RMSE = 11.49 and $R^2$ = 0.92 for the model 1 and RMSE = 7.69 and $R^2$ = 0.96 for the model 2. As the model 2 outperformed the model 1 in both goodness of fit and validation statistics, it was chosen for plot level biomass prediction.

### 5.1.2 Plot-level biomass

Summary of the plot level biomass is shown in Table 4 and a histogram of the plot biomass in Fig. 15. The biomass in the sample plots ranged from 0 to 42 Mg/ha, with mean of 12 Mg/ha. The sample distribution appeared to be slightly right skewed.

*Table 4. Summary of the plot biomass.*

| | Min. | 1st. Quantile | Median | Mean | 3rd. Quantile | Max. |
|---|---|---|---|---|---|---|
| Leaf biomass (Mg/ha) | 0.00 | 5.09 | 10.01 | 12.14 | 15.99 | 42.31 |
| Sisal age | 0 | 3 | 7 | 7.5 | 13 | 17 |

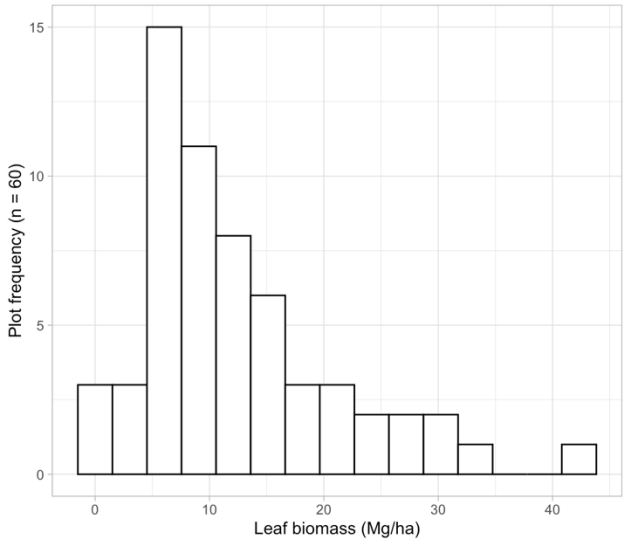


*Fig. 15. Sisal leaf biomass distribution in the plots.*

## 5.2 Remote sensing and biomass modelling

### 5.2.2 Modelling leaf biomass with vegetation indices

The relationship between sisal leaf biomass and individual Sentinel-2 bands is presented in Fig. 16. It shows a relationship that is negative between the biomass and B, G, R, RE1, SWIR1 and SWIR2 bands, and positive between the biomass and RE2, RE3 and NIR bands.

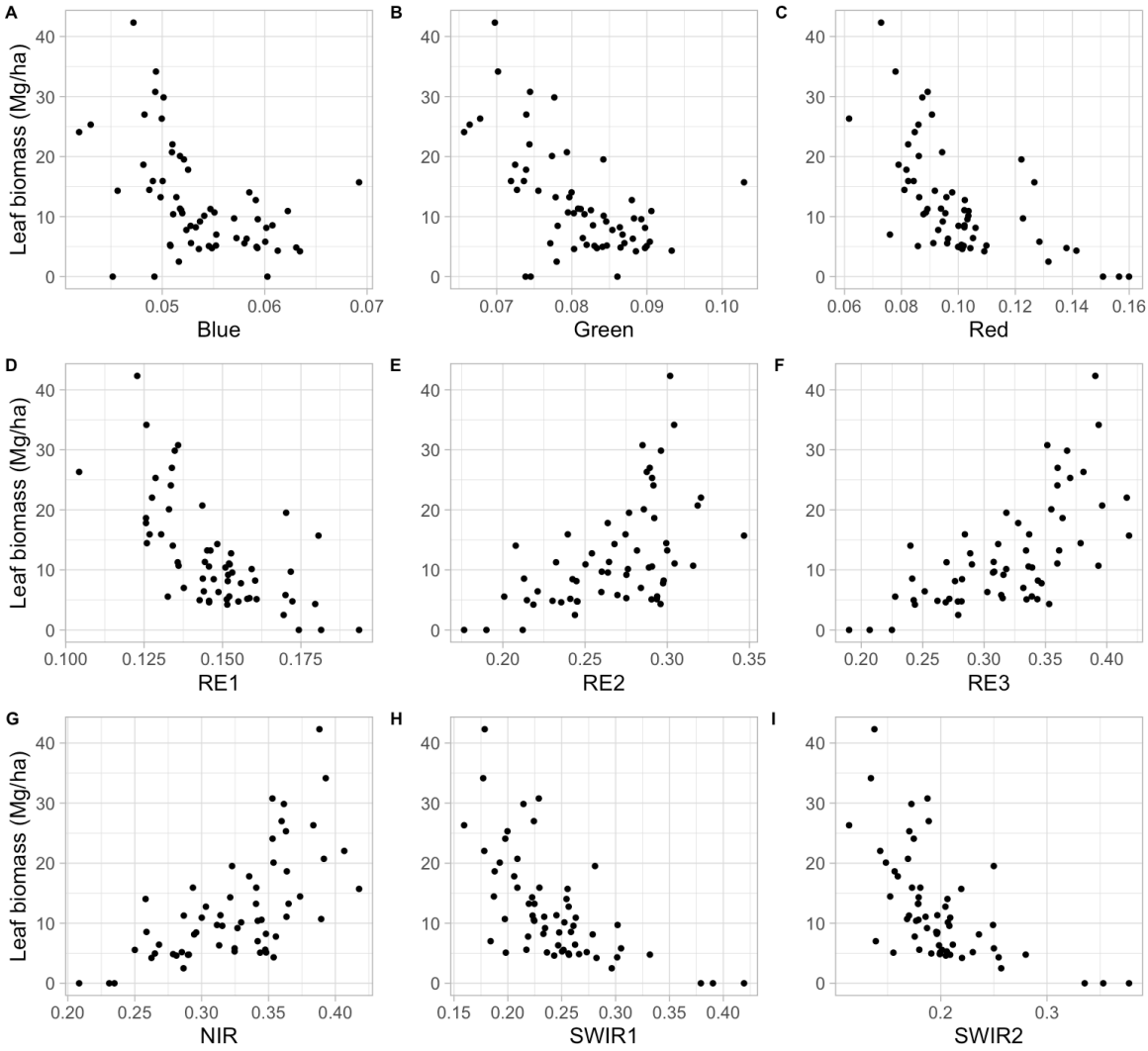


*Fig. 16. The relationship between Sentinel-2 spectral bands and sisal leaf biomass.*

Fig. 17 shows the explained deviances ($D^2$) of the GAMs fitted with all the possible two band combinations for three VI groups (RSI, NDSI, RDSI). In all groups, the best combinations included one of the two red-edge bands (RE2, RE3) or the near-infrared band (NIR), while combinations without these bands showed lower performance. RSI and NDSI results were close to identical and in both groups NIR, RE2, RE3 and B, R, G, RE1 combinations performed better than NIR, RE2, RE3 and SWIR1, SWIR2 combinations. The best combination in these groups was two red-edge bands RE2 and RE3. The RSI with the best explanatory power, which was also best of all the tested VIs was RE2/RE3 ($D^2 = 0.738$), while the NDSI with best explanatory power was (RE3-RE2)/(RE3+RE2) ($D^2 = 0.736$). The combinations of RE3, NIR and G and RE1 such as NIR/G ($D^2 = 0.711$) and (NIR-G)/(NIR+G) ($D^2 = 0.704$) were the second-best choice for RSI and NDSI. Overall, RDSIs showed lower performance than RSIs or NDSIs, but the best band combinations were similar. In RDSI group, the best explanatory power was achieved with 1/Green-1/RE3 ($D^2 = 0.662$) and the second best with 1/RE3-1/RE2 ($D^2 = 0.65$).

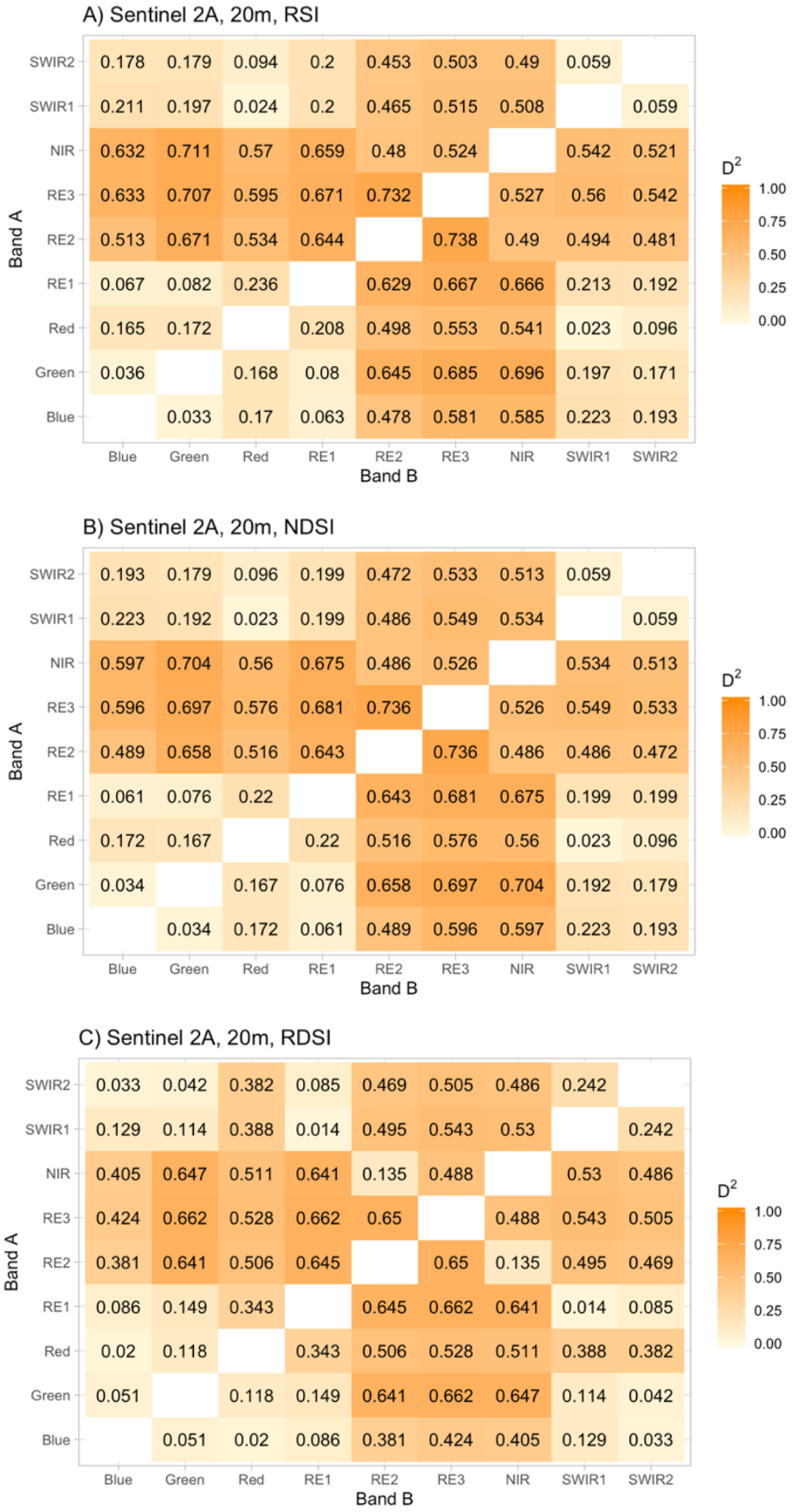


*Fig. 17. Explained deviances ($D^2$) of GAM-models fitted with all the Sentinel-2 two-band combinations of A) ratio based spectral indices (Band A/Band B), B) normalized difference spectral indices ((Band A – Band B)/(Band A + Band B)) and C) reciprocal difference*

GAMs were fitted also for all the reference VIs. These models and the best performing RSI, NDSI and RDSI were cross-validated with the leave-one-out method. The model fit ($D^2$) and validation RMSE of the models are presented in the Table 5. Reference VIs with the best explanatory power were IRECI ($D^2$ = 0.609) and MCARI ($D^2$ = 0.583). The best RSI, NDSI and RDSI, however, outperformed all the reference VIs. Fig. 18 shows best two VIs of all the VI groups and reference VIs, as well as NDVI. Fig. 18, RE2/RE3 is shown as 1–(RE2/RE3) to make the relation to biomass positive, like the other indices.

*Table 5. Explained deviances ($D^2$) and RMSE (root mean squared error) of leave-one-out cross-validation of best RSI, NDSI, RDSI and all the reference VIs.*

| Index | RMSE (Mg/ha) | NMRSE (%) | $D^2$ |
|---|---|---|---|
| RE2/RE3 | 4.959 | 50.2 | 0.738 |
| (RE3-RE2)/(RE3+RE2) | 5.014 | 50.8 | 0.736 |
| 1/G-1/RE3 | 5.630 | 57.0 | 0.662 |
| IRECI | 5.956 | 60.3 | 0.620 |
| MCARI | 5.959 | 60.4 | 0.617 |
| EVI | 5.936 | 60.1 | 0.606 |
| OSAVI | 6.033 | 61.1 | 6.033 |
| CCCI | 6.154 | 62.4 | 0.577 |
| NDVI | 6.234 | 63.2 | 0.560 |
| SR | 6.351 | 64.3 | 0.541 |

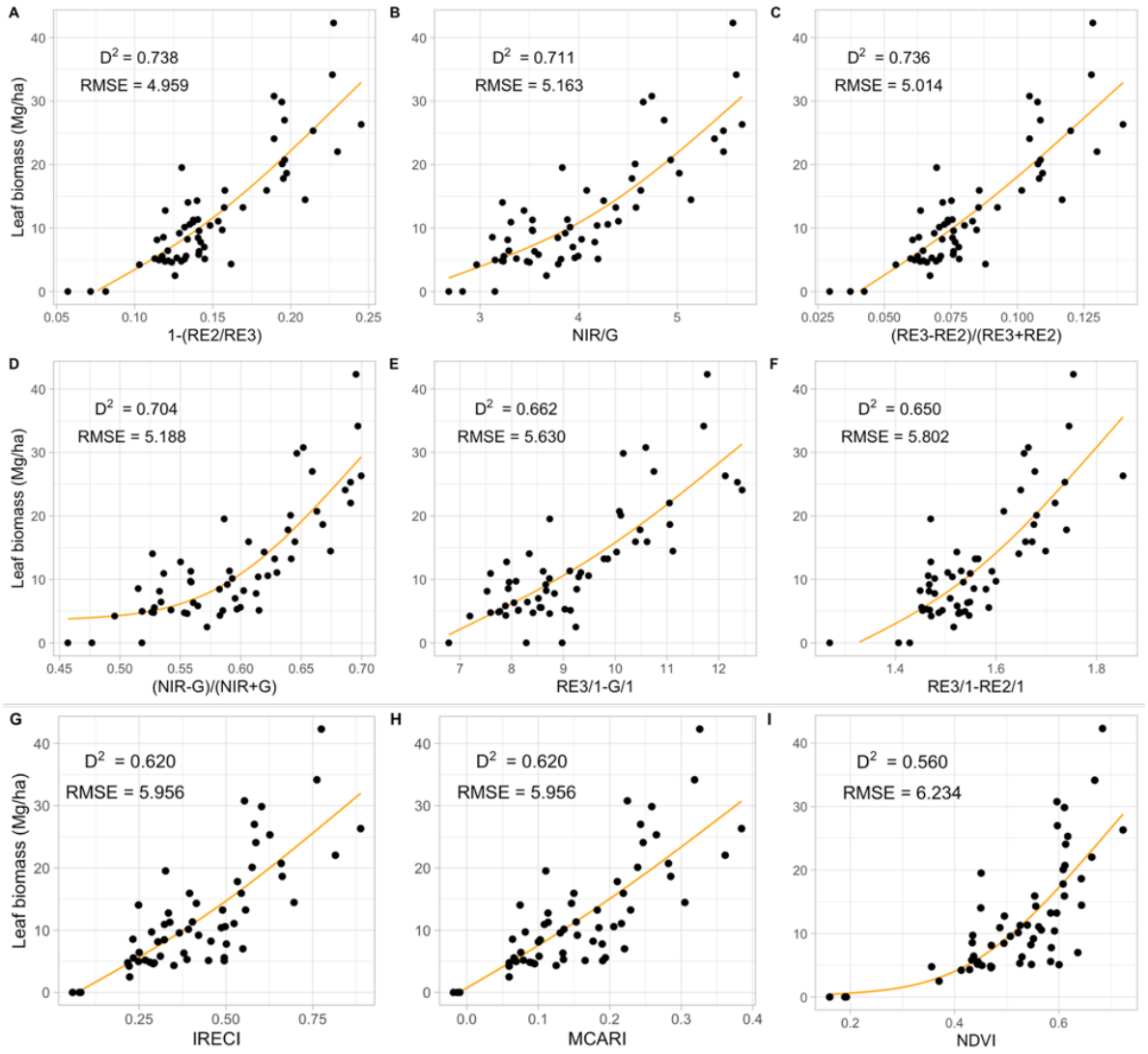


*Fig. 18. Best performing biomass-VI GAMs (orange line). Two VIs from ratio, normalized difference, reciprocal difference and reference groups. Also NDVI as reference. $D^2$ = deviance explained. RMSE = root mean squared error (Mg/ha).*

The impact of the additional plot variables (understory, flower stalks, harvesting status, variety and age) was analysed visually by colouring the best VI-model with each of the variables (Fig. 19). Fig. 19 shows that the highest biomass observations are from the younger fields that have not yet been harvested. On the higher end of the range the understory seems to lead to overestimation of the biomass, although there are a few observations contradicting this trend. The model appears to overestimate biomass also for the youngest fields that have not yet been harvested. The flower stalks, the variety or the age do not seem to introduce notable bias.

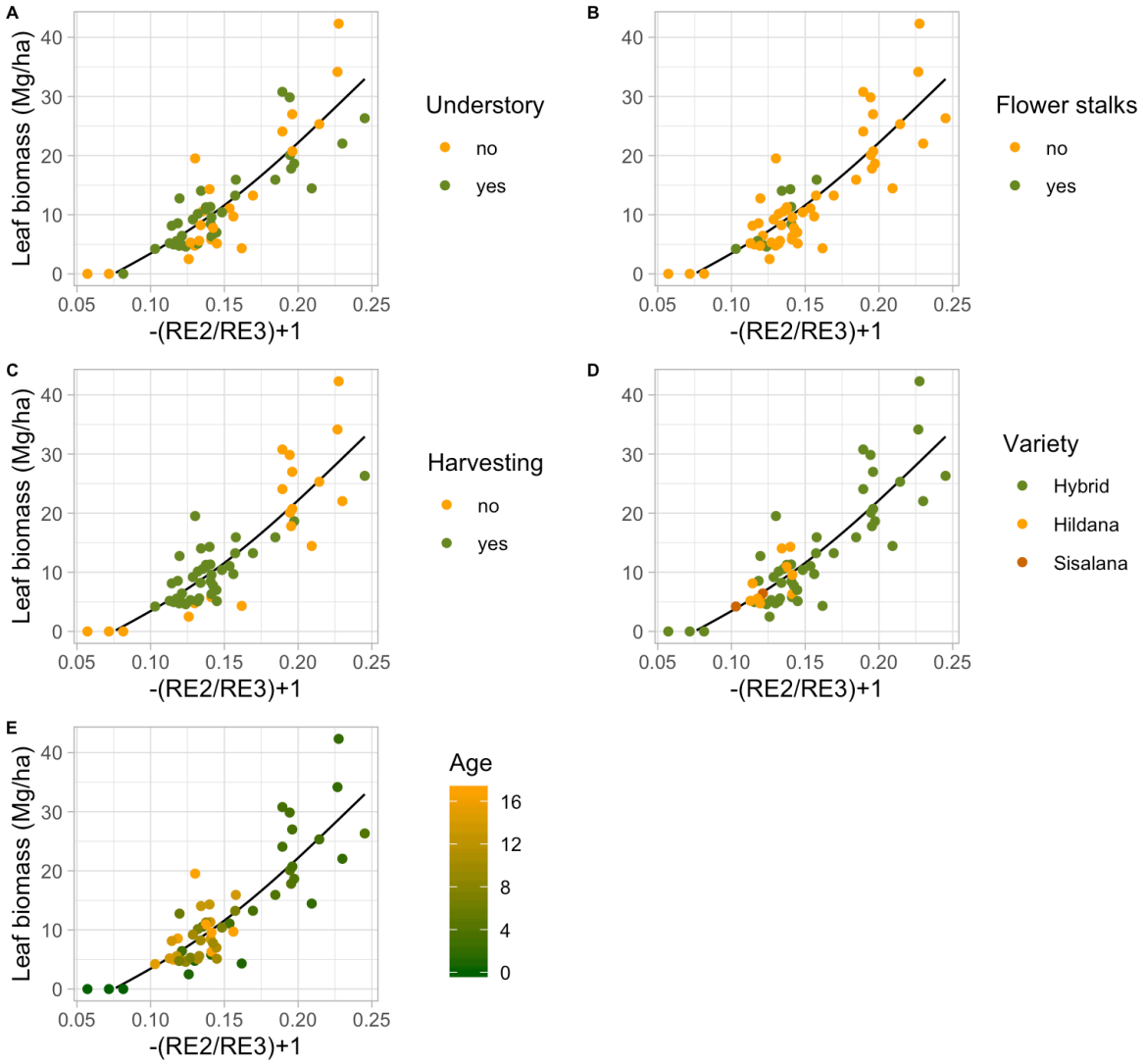


*Fig. 19. The best VI-biomass model, with observations colored based on plot variables.*

### 5.3.2 Leaf biomass map for the study area

Leaf biomass was predicted for the plantation area with the best biomass-VI model (Appendix 2, Fig. 20). Summary of the predicted biomass is shown in Table 6. The biomass in the study area ranged from 0 to 45 Mg/ha. The map shows a high amount of intra-and inter-field variability in the distribution of the biomass. The total amount of the predicted biomass at the plantation was 87 370 Mg.

*Table 6. Summary of the sisal leaf biomass in the study area.*

| | Minimum | 1st. Quantile | Median | Mean | 3rd. Quantile | Maximum |
|---|---|---|---|---|---|---|
| Leaf biomass (Mg/ha) | 0.00 | 5.61 | 8.49 | 9.87 | 12.16 | 45.12 |


*Fig. 20. Sisal leaf biomass in the study area (29 September 2019).*

# 6. DISCUSSION

## 6.1 Sisal leaf allometry

Allometric relationships between plant structure and mass often provide an effective way to estimate biomass, once these relationships are known (Paul *et al*., 2016). The first objective of this research was to develop an allometric model for predicting sisal leaf biomass. It was found that leaf length, maximum diameter and plant height were all highly related to biomass. Combining diameter and height to constitute an approximation of trunk volume is often used in tree allometry to increase the prediction accuracy (Picard *et al*., 2012). Here, the leaf volume was approximated by combining the maximum diameter, with either the length of the leaf, or the height of the plant. A strong linear log-log dependency was found between the leaf volume and its mass. This relationship, which appears to be stable throughout the plant's life cycle, can be expressed with exponential allometric equation which can accurately predict the leaf biomass. Similar relationship between volume and mass has been observed for i.e. tropical trees (Henry *et al*., 2010). For crops, volume is not commonly used as a predictor, but linear log-log dependency between the mass and length, or the mass and diameter, has been noted for maize (Reddy *et al*., 1998).

The volume-mass relationship for sisal leaf was slightly stronger when the volume was approximated using plant height, instead of leaf length. The plant height was measured from topsoil to the highest point of the plant, which is the apex of the unfolding leaf which grows upwards from the middle of the rosette. The leaf length, on the other hand, was measured after the leaves were cut and the cut was done along the narrowest width at the leaf base. It is possible that this resulted in small inconsistencies in length measurements, if the location of the cut was not constant along the leaf axis, because of structural variation or difficulty of cutting leaves from exactly the same spot. This could explain why the use of plant height resulted in slightly more accurate prediction. Nonetheless, the result shows that the mass can also be approximated using the leaf length, but this leads to slightly lower prediction accuracy.

The allometric model was formulated only for the leaves. Hence, to use it to assess the total leaf biomass of a plant, one must also count the number of the leaves of the plant. Therefore, a model for the whole plant would be more practical and would also include the stem biomass. However, the leaves are the part of the plant that is harvested and used, although the stem could also be used for biofuel production (Terrapon-Pfaff *et al*., 2012). Additionally, allometric models for a whole plant are usually formulated and used for intact plants (Sampaio & Silva, 2005). The use of such model could be problematic for perennial crop like sisal, because the

harvesting of the leaves alters the plant structure (i.e. diameter). Then again, including only the leaves means that the model omits some parts of the biomass, mainly the stem. In addition, a small part of the leaf base was omitted when the leaves were cut. These parts of the plant, the stem and the leaf base that is not harvested, are also referred to as sisal ball (see Fig. 7). Dry weight of a fully mature sisal ball is approximated to be 5.8kg (Terrapon *et al*., 2012), which with a plant population of 4995 per ha would add up to 2.90 Mg/ha in a fully mature field. Traditionally, sisal balls are used as manure by burning and ploughing them into the soil, but they could also be used as a feedstock for energy production (Terrapon-Pfaff *et al*., 2012). Future research should quantify the biomass of the sisal ball. Furthermore, from the carbon sequestration perspective, assessing the biomass of the root system would also be highly relevant (Glover 1939, Rasse *et al*., 2005)

The allometric theory states, that the scalable relationships are applicable for plants growing under similar conditions (Niklas, 2004). This notion has been contested by Paul *et al.* (2016), who tested the generality of the AGB-allometry across ecoregions and plant functional types (i.e. shrubs, multi-stemmed trees and trees of different wood densities) at continent scale in Australia. They found that generic models based on plant functional type could be used instead of species-specific models to predict AGB across ecoregions, with only minor losses on prediction accuracy. Conversely, Youkhana *et al.* (2017) observed that rainfall had an impact on AGB growth of tropical perennial grasses and suspected that more biomass was allocated to roots due to lower soil moisture. Presumably, the climatic factors across the sisal plantation are similar, but the management practises such as fertilization and the use of pesticides introduce heterogeneity to growing conditions. Samples of the allometric model had grown under variety of management practices, but this did not seem to have an impact on the found relation. Also, the sisal variety had no impact. This suggest that the model is applicable regardless of the variety or management practises. The effect of climate factors should however be studied, because sisal is grown in both tropical and subtropical regions, in countries with different climate types (FAO, 2020).

## 6.2 Modelling sisal leaf biomass with multispectral satellite imagery

Remote sensing is a feasible method for a large-scale assessment and monitoring of crop biomass, but the methods and outcomes are often species specific (Ahamed *et al*., 2011). The second objective of this thesis was to assess the utility of medium-resolution multispectral satellite imagery in estimating the leaf biomass of sisal. This objective was pursued by exploring the relationships between Sentinel-2 VIs and the leaf biomass. The 8851 ha study area consists of fields at various growing stages and under different management practises. The

results show a strong relationship between multispectral vegetation indices and sisal leaf biomass. The best results were achieved with the ratios of RE2 to RE3 and NIR to G bands, which outperformed all the reference indices, and other band combinations.

Out of the three VI groups that were tested, ratio indices and normalized difference indices had overall stronger relation to biomass than reciprocal difference indices, or reference indices. The most valuable single bands for the biomass assessment were found to be RE2, RE3 and NIR. Ratios including at least one of these bands resulted in satisfactory model fit, while ratios without these bands had low performance. VI with the strongest relation to the leaf biomass was RE2/RE3, while second-best ratio VI was NIR/G. Almost identical relations were found using the same bands, but the normalized difference formula. The two best ratio indices appear to be similar to Red-edge Chlorophyll Index ($CI_{Red\text{-}edge}$ = NIR/RE) and to Green Chlorophyll Index ($CI_{Green}$ = NIR/G), while the normalized difference ratio indices are similar to Green NDVI (GNDVI = (NIR-G)/(NIR+G)) and Red-Edge NDVI ($NDVI_{re}$ = (NIR-RE)/(NIR+RE)), all of which Gitelson and Merzlyak (1994) and Gitelson *et al.* (1996, 2003) have found to be sensitive to LAI and green leaf biomass at leaf and canopy level. The applicability of these indices for crop studies, using Sentinel-2 data, have also been demonstrated before (J. G.P.W. Clevers & Gitelson, 2013; Viña *et al*., 2011).

The distinctive feature of the red-edge indices, which showed the highest sensitivity to sisal leaf biomass, is that even though they are broadly based on the same spectral regions as in previous crop studies, they were calculated using the bands with slightly different positioning. Originally, $CI_{Red\text{-}edge}$ and and $NDVI_{re}$ were formulated using narrow spectral bands (1 nm) as 750/700 nm and (750–700 nm)/(750–700 nm), respectively (Gitelson *et al*., 2003; Gitelson & Merzlyak, 1994). Hence, when calculated from multispectral data the band corresponding to NIR has usually been used as numerator and the band nearest to 700 nm as denominator. For example Kross *et al.* (2015) calculated $NDVI_{re}$ using RapidEye NIR (760–850 nm) and RE (690–730 nm) bands, when estimating the LAI and biomass of corn and soybean. From Sentinel-2 data, Clevers *et al*. (2017) calculated $CI_{Red\text{-}edge}$ using RE3 (773–793 nm) and RE1 (698–713 nm) for retrieving chlorophyll and LAI of potato. Here, the best performing red-edge indices were calculated using RE2 (733–748 nm) and RE3 (773–793 nm) bands. Using RE2, instead of the band closest to 700 nm as in other crop studies, increased the explained deviance by 9.3% at best.

The functioning of the red-edge indices is based on the notion that the red-edge position (point of the maximum slope between R and NIR wavelengths) is mainly controlled by chlorophyll

concentration, shifting to higher wavelengths with increasing chlorophyll (Horler *et al*., 1983). Therefore it can provide information about various plant parameters. It has been found that reflectance near 700 nm is a sensitive indicator of the red-edge position, and furthermore, that the ratio of 750 nm to that near 700 nm is directly proportional to chlorophyll content, as well as green leaf biomass (Gitelson *et al.*, 1996, 2003). However, these relations were observed for maize with leaf biomass of 0 to 3.5 Mg/ha, while sisal leaf biomass in the study area ranged from 0 to 42.1 Mg/ha, with the mean of 9.9 Mg/ha. Needless to say, also the structural differences at leaf and canopy level between these plants are obvious. Presumably, as a result of higher biomass quantities and the consequent expansion of chlorophyll concentration, sisal red-edge position can move to longer wavelengths than that of maize. This is supported with two notions. The first indication of it is that the RE1 band, centred at 705 nm, had a negative relation to biomass, resembling the spectral response of the red band. Secondly, RE2/RE3 (733–748 nm/773–793 nm), the ratio that was most sensitive to biomass, was found from longer wavelengths.

Relatively high biomass values and the structure of the plant are likely to explain why reference VIs, all of which included the R band, had weaker sensitivity to biomass. Although NIR/red ratios are known to correlate with biomass, they tend to saturate at high values (Holben *et al*., 1980). This is because at red wavelengths, the absorption coefficient of chlorophylls is high and the depth of light penetration into the leaf is low (Kumar & Silva, 1973; Lichtenthaler, 1987). Sisal, on the other hand, has high amount of green biomass and consequently high chlorophyll concentration, thick leaves and multiple leaf layers. This could explain why the R band showed low sensitivity to the biomass. In addition to the R band, the best reference VI, IRECI, contained also the RE bands, whereas the reference VIs based solely on the NIR and R bands, like SR, showed lower performance. The usefulness of the RE bands over the R band is clearly demonstrated also by comparing $NDVI_{re}$ – which was substantially better indicator of the biomass – to NDVI, which explained 18% less of the deviance.

In addition to the NIR and RE bands, also the G band was found to be sensitive to the biomass. A strong relation, almost as strong as for RE-indices, was found for NIR/G and (NIR-G)/(NIR+G), known as $Cl_{green}$ and GNDVI. These indices appear also to be slightly more sensitive to lower biomass values than the RE indices. Just like the red-edge position, the reflectance at the green spectral region is controlled by the chlorophyll concentration (Gitelson and Merzlyak, 1994). In this region, the absorption coefficient of chlorophyll is smaller, and light can penetrate deeper into the leaf, than at the red region (Kumar & Silva, 1973; Lichtenthaler, 1987). Therefore, the green region does not saturate as easily, and is highly

sensitive to changes in chlorophyll concentration (Gitelson and Merzlyak, 1994; Gitelson *et al*., 1996). Although the NIR/R combinations have been the go-to-option in vegetation studies, Gitelson *et al*. (1996) have advocated the use of NIR/G combinations, because of the wider dynamic range of the G bands. In this study, the G band showed higher sensitivity to biomass than the R band, with deviance explained by NIR/G clearly higher than by NIR/R.

One challenge of the VI-modelling approach is how to minimize the impact of the external factors and make the relationship to biomass most sensitive (Chao *et al*., 2019). The performance of the best models appears to be slightly lower than in other studies that have used multispectral data for crop biomass assessment. For example, Kross *et al.*, (2015) achieved $R^2$ ranging between 0.86 and 0.88 when mapping corn and soybean biomass with RapidEye imagery. With handheld multispectral spectrometer, Prabhakara *et al.* (2015) achieved $R^2$ of 0.86 for six winter crops. For wheat biomass assessment, Wang *et al.* (2016) used HJ1-satellite, which resulted in $R^2$ of 0.79. Then again, for biomass assessment of rangeland grasses with Sentinel-2, Sibanda *et al*. (2015) achieved slightly lower $R^2$ of 0.58.

Compared to the crops in the above-mentioned studies (i.e. soybean, corn, wheat), which grow in more homogeneous fields, the growing conditions of sisal in the study area are rather different. Although it is a monoculture plantation, the varying management practises have resulted in fields that are heterogeneous in terms of the understory between the sisal rows. Some fields have no, or just little weeds, while in the other fields the 3.75m space between the rows is fully covered with tall grass or shrubs, or both. With Sentinel's 20m resolution, this means that the signal is always a mixture of sisal and soil or understory, or both. Spectral regions used in the red-edge and NIR to G ratios, which showed highest sensitivity to biomass, have shown low sensitivity to soil background (Prudnikova *et al*., 2019; Viña *et al*., 2011). Yet it should be acknowledged, that the VIs sensitivity to soil background depends on the soil type (Prudnikova *et al*., 2019). Understory, however, is more likely to introduce some uncertainty to the models. 19A, which shows the RE2/RE3 model with colouring based on presence/absence of understory, provides some means to assess this. The trend, although not fully consistent, seems to be that for lower biomass values the effect of understory is minor. For higher values, the presence of understory seems to lead to overestimation of biomass. The model appears to be overestimating the biomass also for 1–2-year-old fields, which have not yet been harvested.

The acquisition date of the Sentinel-2 image was a month apart from the fieldwork dates. Therefore, any changes in the field locations due to management practices and harvesting cannot be completely ruled out. The changes in the understory vegetation are supposedly small,

as both the field campaign and the image acquisition dates fall into the middle of the dry season. During a year, however, the seasonality of the rains (Fig. 4) and resulting phenological patterns cause the understory to change from dry during the dry seasons to green during the rainy seasons. If the understory causes bias as assumed, the models potentially perform worse during the rainy seasons, due to the greening of the herbaceous vegetation (Horion *et al*., 2014). Potential sources of error include also positional inconsistencies between the field plots and the image. Although they are likely to be minor, considering the plot size and the precision with which the locations were measured. Error from the allometric model, which showed a good predicting capacity, is also likely to be small. Measuring errors during the fieldwork, on the other hand, can have occurred.

The one-month gap, between the field work dates and the acquisition date of the nearest fully cloud free Sentinel-2 image of the study area, demonstrates one of the shortcomings of the optical satellites. In areas with regular cloud cover the potential benefits of multispectral satellites' frequent revisit time (less than one week for Sentinel-2) can seldom be realized (Asner, 2001). This makes the availability of the data unpredictable and hampers the possibility of frequent monitoring. Future research should address the use of complementary data sources, such as SAR, which is not affected by cloud cover, but can also be used for crop monitoring (Wiseman *et al.*, 2014). Combined with multispectral data, in a fusion model, SAR could also enhance the model performance (Chang & Shoshany, 2016). Also UAVs are a potential complementary or alternative data source for biomass modelling (Han *et al*., 2019). Furthermore, the interactions with the understory seemed to introduce bias into the models and hence require further studies. Considering especially the temporal transferability of the model, which can be affected by the phenological patterns of the understory (Horion *et al.*, 2014). Accounting for seasonal features could enhance the model performance (Liu *et al*. 2016). Also, with UAVs or other data sources with comparable spatial resolution, the background effect of the understory could probably be diminished, since earlier research has demonstrated the possibility of segmenting *Agave* crop rows from fine resolution data (Calvario et al., 2017). Efforts on minimizing the external factors would also benefit the spatial transferability of the model, which adds to the list of future inquiries.

To summarise, a strong relation was found between sisal leaf biomass and multispectral VIs. Best performing VI was the ratio of RE2 (733–738 nm) to RE3 (773–793 nm), which resembles the $Cl_{rededge}$ (A. Gitelson & Merzlyak, 1994). However, with a distinction that in addition to a band in the NIR region, it uses a band centred at 740 nm instead of 700 nm. This combination do not seem to have been used in earlier multispectral crop studies, which have calculated

$CI_{rededge}$ using the original formula (i.e. in Frampton *et al.* (2013) and in Clevers *et al*. (2017)). This result underlines the value of the red-edge bands in biomass assessment and also that the positioning of the red-edge bands matters. It also shows that to take full advantage of the red-edge bands all the possible band combinations should be tested, instead of only the conventional formulas. In addition to the red-edge ratios, a strong relation to biomass was found from NIR (855–875 nm) to G (543–578 nm) ratio, known as $CI_{green}$ (A. Gitelson & Merzlyak, 1994). The advantage of this VI, as noted by Clevers *et al*. (2017), is that it avoids the need for red-edge bands, which in Sentinel-2 are available only at 20m resolution. Hence it can be calculated also with 10m resolution, although this does not necessarily boost the model performance (Riihimäki *et al*., 2019). Furthermore, unlike the red-edge bands, the NIR and G bands are found from all the multispectral satellites, which makes this combination more available and interoperable with other optical sensors.

## 6.3 Sisal leaf biomass in the study area

The final objective of this thesis was to model sisal leaf biomass at plantation level. This was done using the best performing VI-biomass model (RE2/RE3). The result is a 20m resolution map, which shows the distribution of the leaf biomass across the plantation. The highest biomass densities were found from 2–5-year-old fields, where the harvesting has not yet begun, or has just started. Fields younger and older than this had lower amounts of biomass. The distribution of the leaf biomass is mainly controlled by the plant age and harvesting status. Due to the repeated harvesting, the biomass does not increase linearly with age, but the lowest biomass densities are found from the oldest and freshly cultivated fields. After planting, the biomass increases rapidly during the first years, and reaches the maximum density in 2-5 years. At this age the regular harvesting of the leaves begins, decreasing the biomass. As a result, it is mainly the fields older than this where the biomass is close to the mean. When the harvesting and the plants life cycle approaches the end, the biomass is again at the lower end of the range.

The plantation consists of field blocks, all at the different stage of the growing cycle. The resulted map shows that at 20m resolution variations in the distribution of the biomass can be detected across the plantation, but also inside the field blocks. Many of the fields appear internally heterogeneous with randomly distributed spots of low biomass, and spatial patterns indicating more rigorous growth in some parts of the fields. In addition, during the field work it was observed that in the older fields some of the plants have been fully harvested, while others still have leaves. In the map these areas exhibit irregular patterns. These notions suggest that the resolution would be sufficient for monitoring the growth and detecting disturbances in

the individual field blocks. By using the NIR to G ratio, the map can be produced at 10m resolution, adding even more detail to the distribution of the biomass.

Adjacent of the study area, in the lowlands near the Taita Hills, the main land cover types are cropland, grassland and Acacia-Commiphora bushland, of which the latter is closest to natural vegetation in the area (Abera *et al*., 2020; Pellikka *et al.*, 2018; Pellikka *et al.*, 2013). Woody AGB of these land cover classes has been assessed by Pellikka *et al.* (2018) and Amara *et al*. (in prep). In Pellikka *et al.* (2018) the bushland was further divided into shrubland and thicket. The mean sisal leaf biomass at the plantation (9.87 Mg/ha) corresponds to the AGB of bushland (9.0 Mg/ha), or shrubland (6.80 Mg/ha) and thicket (11.60 Mg/ha). In the croplands, the mean AGB (4.91 Mg/ha in Pellikka *et al*. (2018) and 5.80 Mg/ha in Amara *et al*. (in prep.)) is lower than the mean leaf biomass at the plantation. However, the internal variation of biomass density at the plantation is high. The lowest densities are at the level of the grasslands (1.8 Mg/ha) (Amara *et al*, in prep.), while the largest densities are similar to plantation forests (43 Mg/ha) (*Pellikka et al.,* 2018). The temporal changes of the biomass at the plantation require further investigation, but the mean value and hence the total amount of stored carbon is likely to be stable over time due to the rotational harvesting, assuming the production rates remain steady. Thus the carbon stored in the leaf biomass at the plantation appears to be at the same level as in the woody AGB of the bushlands, and higher than in the woody AGB of the other croplands.

For a comprehensive understanding of the carbon cycle at the plantation, crop and soil carbon fluxes and the belowground carbon dynamics must be considered (Revill *et al.*, 2013; Ferchaud *et al.*, 2016). Future research should be done on the carbon fluxes, as well as soil carbon sequestration, which in croplands is mainly controlled by the input from the root system, management practices and the soil type (Anderson-Teixeira *et al*., 2013; Sartori *et al*., 2006). The results of this thesis will benefit these future research efforts.

# 7. CONCLUSIONS

In this thesis, the leaf biomass of *Agave sisalana* was assessed at different scales. First, at leaf level by sampling and measuring the leaves. Then, at plantation level using field measurements and Sentinel-2 multispectral satellite data with 20m spatial resolution. There were three objectives: (1) to develop an allometric biomass equation for sisal leaves, (2) assess the utility of medium-resolution multispectral satellite imagery in estimating sisal leaf biomass, and (3) to model the leaf biomass at plantation level.

A strong log-log linear relationship was found between the leaf biomass and volume, which was approximated using the leaf maximum diameter and plant height. A strong relationship was also found between multispectral vegetation indices and sisal leaf biomass. The highest performance was achieved with vegetation indices based on the red-edge ($R_{740}$ and $R_{783}$), near-infrared ($R_{865}$) and green ($R_{560}$) regions. Finally, a biomass map was produced, which showed the distribution of the leaf biomass at the plantation and revealed that the biomass ranged from 0 to 45.1 Mg/ha, with mean at 9.9 Mg/ha. The largest biomass densities were found from 2-4-old fields, which have not yet been harvested, while freshly cultivated fields and the oldest fields had the lowest biomass densities.

Because the allometric model included only the leaves, the allometry of the omitted plant parts – the stem, the flower stalk and the roots – should be studied in future. The performance of the VI-biomass models was slightly lower than in previous crop studies. The main factor reducing the model performance appears to be the heterogeneity of the understory vegetation. Another limitation of this modelling approach is the irregular availability of the data due to cloud cover potential. The model performance could be improved by studying the background effect and incorporating other data sources into the model. Testing of complementary data sources are recommend also to counter data availability issue. The biomass modelling at plantation level showed that the mean leaf biomass and hence the stored carbon is comparable to the semi-natural *Acacia-Commiphora* bushlands. For a complete understanding of the carbon cycle at sisal plantation, soil and plant carbon fluxes and soil carbon sequestration have to be quantified.

To the authors best knowledge, this was the first study that assessed the biomass of *Agave sisalana* or any other *Agave*-species using allometric modelling and remote sensing. The leaf modelling resulted in an allometric equation that can be used as an accurate tool for non-destructive leaf biomass assessment. The VI-biomass modelling results showed that multispectral data is suitable for assessing the biomass at plantation level and in individual field blocks, but the model performance is limited by the background effect of the understory.

Nevertheless, the resulted biomass model will benefit future analyses of plant productivity and carbon sequestration.

# 8. AKNOWLEDGEMENTS

Funding from the of Academy Finland is gratefully acknowledged for SMARTLAND (Environmental sensing of ecosystem services for developing a climate-smart landscape framework to improve food security in East Africa, decision no. 318645).

I sincerely thank Teita Sisal Estate management, namely Mr. Emmanuel Mrombo, for a generous cooperation throughout the research. Taita Taveta University supported the research by arranging the use of their laboratory, for which I am deeply grateful. Thank you also Mr. Dickson Wachira for your kind assistance in the laboratory. Essential for the completion of this thesis were the working facilities at the Taita Research Station in Wundanyi, Kenya and at the department of Geosciences and Geography in the University of Helsinki, Finland. Thank you Mr. Mwadime Mjomba from the Taita Research Station for helping to organize the field work and assisting during it. Thank you also Mr. Darius Kimuzi from the Taita Research Station for participating in the field work. I also extend my thanks to everyone else working at the Taita Research Station. Last but not least, I wish to express my honest gratitude and respect for my supervisors. Thank you Professor Petri Pellikka who originally proposed me this thesis topic, helped to organize the research and was greatly supportive all through the process. Special thanks to Dr. Janne Heiskanen who guided the research and participated in the field work, thus contributing significantly to a successful completion of this thesis.

# 9. REFERENCES


Abera, T. A., Heiskanen, J., Pellikka, P. K. E., Adhikari, H., & Maeda, E. E. (2020). Climatic impacts of bushland to cropland conversion in Eastern Africa. *Science of the Total Environment*, 717, 137255. https://doi.org/10.1016/j.scitotenv.2020.137255

Ahamed, T., Tian, L., Zhang, Y., & Ting, K. C. (2011). A review of remote sensing methods for biomass feedstock production. *Biomass and Bioenergy*, *35*(7), 2455–2469. https://doi.org/10.1016/J.BIOMBIOE.2011.02.028

Anderson-Teixeira, K. J., Masters, M. D., Black, C. K., Zeri, M., Hussain, M. Z., Bernacchi, C. J., & DeLucia, E. H. (2013). Altered Belowground Carbon Cycling Following Land-Use Change to Perennial Bioenergy Crops. *Ecosystems,* 16, 508-520. https://doi.org/10.1007/s10021-012-9628-x

Asner, G. P. (2001). Cloud cover in Landsat observations of the Brazilian Amazon. *International Journal of Remote Sensing*, 22(18), 3855-3862. https://doi.org/10.1080/01431160010006926

Barnes, E. M., Clarke, T. R., Richards, S. E., Colaizzi, P. D., Haberland, J., Kostrzewski, M., Waller, P., Choi C., R. E., Thompson, T., Lascano, R. J., Li, H., & Moran, M. S. (2000). Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. *Proc. 5th Int. Conf. Precis Agric*.

Baskerville, G. L. (1974). Use of Logarithmic Regression in the Estimation of Plant Biomass: Reply. *Canadian Journal of Forest Research,* 4(1), 149. https://doi.org/10.1139/x74-024

Battude, M., Al Bitar, A., Morin, D., Cros, J., Huc, M., Marais Sicre, C., Le Dantec, V., & Demarez, V. (2016). Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data. *Remote Sensing of Environment,* 184, 668-681. https://doi.org/10.1016/j.rse.2016.07.030

Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., Gnyp, M. L., & Bareth, G. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *International Journal of Applied Earth Observation and Geoinformation*, 39, 79-87. https://doi.org/10.1016/j.jag.2015.02.012

Birth, G. S., & McVey, G. R. (1968). Measuring the Color of Growing Turf with a Reflectance Spectrophotometer 1 . *Agronomy Journal*, 60(6). https://doi.org/10.2134/agronj1968.00021962006000060016x

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, 47(5), 1287-1294. https://doi.org/10.2307/1911963

Calvario, G., Sierra, B., Alarcón, T. E., Hernandez, C., & Dalmau, O. (2017). A multi-disciplinary approach to remote sensing through low-cost UAVs. *Sensors,* 17(6), 1411. https://doi.org/10.3390/s17061411

Castillo, J. A. A., Apan, A. A., Maraseni, T. N., & Salmo, S. G. (2017). Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the Philippines using Sentinel imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 134, 70-85. https://doi.org/10.1016/j.isprsjprs.2017.10.016

Chang, J., & Shoshany, M. (2016). Mediterranean shrublands biomass estimation using Sentinel-1 and Sentinel-2. *International Geoscience and Remote Sensing Symposium (IGARSS)*. https://doi.org/10.1109/IGARSS.2016.7730380

Chao, Z., Liu, N., Zhang, P., Ying, T., & Song, K. (2019). Estimation methods developing with remote sensing information for energy crop biomass: A comparative review. *Biomass and Bioenergy*, *122*, 414–425. https://doi.org/10.1016/J.BIOMBIOE.2019.02.002

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B. C., Duque, A., Eid, T., Fearnside, P. M., Goodman, R. C., Henry, M., Martínez-Yrízar, A., Mugasha, W. A., Muller-Landau, H. C., Mencuccini, M., Nelson, B. W., Ngomanda, A., Nogueira, E. M., Ortiz-Malavassi, E., Pélissier, R., Ploton, P., Ryan, C. M., Saldarriaga, J. G., & Vieilledent, G. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology*, *20*(10), 3177–3190. https://doi.org/10.1111/gcb.12629

Claverie, M., Demarez, V., Duchemin, B., Hagolle, O., Ducrot, D., Marais-Sicre, C., Dejoux, J. F., Huc, M., Keravec, P., Béziat, P., Fieuzal, R., Ceschia, E., & Dedieu, G. (2012). Maize and sunflower biomass estimation in southwest France using high spatial and temporal resolution remote sensing data. *Remote Sensing of Environment,* 124, 844-857. https://doi.org/10.1016/j.rse.2012.04.005

Clevers, J. G.P.W., & Gitelson, A. A. (2013). Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on sentinel-2 and-3. *International Journal of Applied Earth Observation and Geoinformation,* 23, 344-351. https://doi.org/10.1016/j.jag.2012.10.008

Clevers, Jan G.P.W., Kooistra, L., & van den Brande, M. M. M. (2017). Using Sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop. *Remote Sensing,* 9(5), 405. https://doi.org/10.3390/rs9050405

Daughtry, C. S. T., Walthall, C. L., Kim, M. S., De Colstoun, E. B., & McMurtrey, J. E. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment,* 74(2), 229-239. https://doi.org/10.1016/S0034-4257(00)00113-9

Davis, S. C., Dohleman, F. G., & long, S. P. (2011). The global potential for Agave as a biofuel feedstock. *GCB Bioenergy*, *3*(1), 68–78. https://doi.org/doi:10.1111/j.1757-1707.2010.01077.x

Debnath, M., Pandey, M., Sharma, R., S. Thankur, G., & Lal, P. (2010). Biotechnological intervention of Agave sisalana: A unique fibe yielding plant with medicinal property. *Journal of Medicinal Plant Research*, 4(3), 177-187.

Eamus, D., Huete, A., & Yu, Q. (2016). *Vegetation Dynamics*. Cambridge University Press. https://doi.org/10.1017/cbo9781107286221

Escamilla-Treviño, L. L. (2012). Potential of Plants from the Genus Agave as Bioenergy Crops. *BioEnergy Research*, *5*(1), 1–9. https://doi.org/10.1007/s12155-011-9159-x

FAO (2020). Food and Agriculture Organisation of the United Nations, *FAOSTATS,* Data: Crops. Retrieved 1 April 2020 from http://www.fao.org/faostat/en/#data/QC

Ferchaud, F., Vitte, G., & Mary, B. (2016). Changes in soil carbon stocks under perennial and annual bioenergy crops. *GCB Bioenergy*, 8(4). https://doi.org/10.1111/gcbb.12249

Field, C. B., Campbell, J. E., & Lobell, D. B. (2008). Biomass energy: the scale of the potential resource. *Trends in Ecology and Evolution*, 23(2), 65-72. https://doi.org/10.1016/j.tree.2007.12.001

Filella, I., & Peñuelas, J. (1994). The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *International Journal of Remote Sensing*, 15(7), 1459-1470. https://doi.org/10.1080/01431169408954177

Foody, G. M., Boyd, D. S., & Cutler, M. E. J. (2003). Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remote Sensing of Environment* 85(4), 463-474.. https://doi.org/10.1016/S0034-4257(03)00039-7

Forkuor, G., Dimobe, K., Serme, I., & Tondoh, J. E. (2018). Landsat-8 vs. Sentinel-2: examining the added value of sentinel-2's red-edge bands to land-use and land-cover mapping in Burkina Faso. *GIScience and Remote Sensing*, 55(3), 331-354. https://doi.org/10.1080/15481603.2017.1370169

Frampton, W. J., Dash, J., Watmough, G., & Milton, E. J. (2013). Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 82, 83-92. https://doi.org/10.1016/j.isprsjprs.2013.04.007

Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS- MODIS. *Remote Sensing of Environment*, 58(3), 289-298. https://doi.org/10.1016/S0034-4257(96)00072-7

Gitelson, A. A., Keydan, G. P., & Merzlyak, M. N. (2006). Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Research Letters*, 33(11). https://doi.org/10.1029/2006GL026457

Gitelson, A. A., & Merzlyak, M. N. (1996). Signature analysis of leaf reflectance spectra: Algorithm development for remote sensing of chlorophyll. *Journal of Plant Physiology*, 148(3-4), 494-500. https://doi.org/10.1016/S0176-1617(96)80284-7

Gitelson, A. A., Merzlyak, M. N., & Lichtenthaler, H. K. (1996). Detection of red edge position and chlorophyll content by reflectance measurements near 700 nm. *Journal of Plant Physiology*, 148(3-4), 501-508. https://doi.org/10.1016/S0176-1617(96)80285-9

Gitelson, A. A., Viña, A., Arkebauer, T. J., Rundquist, D. C., Keydan, G., & Leavitt, B. (2003). Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophysical Research Letters*, 30(5). https://doi.org/10.1029/2002gl016450

Gitelson, A., & Merzlyak, M. N. (1994). Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus hippocastanum L. and Acer platanoides L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. *Journal of Plant Physiology*, 143(3), 286-292. https://doi.org/10.1016/S0176-1617(11)81633-0

Gnyp, M. L., Bareth, G., Li, F., Lenz-Wiedemann, V. I. S., Koppe, W., Miao, Y., Hennig, S. D., Jia, L., Laudien, R., Chen, X., & Zhang, F. (2014). Development and implementation of a multiscale biomass model using hyperspectral vegetation indices for winter wheat in the North China Plain. *International Journal of Applied Earth Observation and Geoinformation*, 33, 232-242. https://doi.org/10.1016/j.jag.2014.05.006

Goswami, S., Gamon, J. A., Vargas, S., Tweedie, C. E. (2015). Relationships of NDVI, Biomass, and Leaf Area Index (LAI) for six key plant species in Barrow, Alaska. *PeerJ Preprints*. https://doi.org/10.7287/peerj.preprints.913v1

Gould, W. (2000). Remote sensing of vegetation, plant species richness, and regional biodiversity hotspots. *Ecological Applications*, 10(6). https://doi.org/10.1890/1051-0761(2000)010[1861:RSOVPS]2.0.CO;2

Han, L., Yang, G., Dai, H., Xu, B., Yang, H., Feng, H., Li, Z., & Yang, X. (2019). Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. *Plant Methods*, *15*(1). https://doi.org/10.1186/s13007-019-0394-z

Helsen K., Van Cleemput E., Bassi L., Somers B. and Honnay O.. (2020). Leaf spectra of 36 species growing in Rosa rugosa invaded coastal grassland communities in Belgium. Data set. Retrieved 6 June, 2020, from the Ecological Spectral Information System: https://ecosis.org/package/leaf-spectra-of-36-species-growing-in-rosa-rugosa-invaded-coastal-grassland-communities-in-belgium

Henry, M., Besnard, A., Asante, W. A., Eshun, J., Adu-Bredu, S., Valentini, R., Bernoux, M., & Saint-André, L. (2010). Wood density, phytomass variations within and among trees, and allometric equations in a tropical rainforest of Africa. *Forest Ecology and Management*, 260(8), 1375-1388. https://doi.org/10.1016/j.foreco.2010.07.040

Holben, B. N., Tucker, C. J., & Fan, C. J. (1980). Spectral assessment of soybean leaf area and lear biomass. *Photogrammetric Engineering & Remote Sensing*, 46(5), 561-656.

Horion, S., Fensholt, R., Tagesson, T., & Ehammer, A. (2014). Using earth observation-based dry season NDVI trends for assessment of changes in tree cover in the Sahel. *International Journal of Remote Sensing*, 35, 2493-2515. https://doi.org/10.1080/01431161.2014.883104

Horler, D. N. H., Dockray, M., & Barber, J. (1983). The red edge of plant leaf reflectance. *International Journal of Remote Sensing*, 4(2), 273-288. https://doi.org/10.1080/01431168308948546

Huete, A., Justice, C., & Leeuwen, W. van. (1999). MODIS Vegetation Index (MOD 13). In *Algorithm Theoretical Basis Document*. (ATBD) Version 3.0. EOS Project Office, NASA Goddard Space Flight Center, Greenbelt, MD (1999), p. 2

IPCC. (2006). *2006 IPCC Guidelines for National Greenhouse Gas Inventories: Agriculture, Forestry and Other Land Use*, Chapter 4.

IPCC. (2019). *Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems.*

Jensen, J. R. (2014). *Remote sensing of the environment: an earth resource perspective.* Second edition. Pearson Education Limited, Harlow, England.

Kell, D. B. (2012). Large-scale sequestration of atmospheric carbon via plant roots in natural and agricultural ecosystems: Why and how. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 356, 1589–1597. https://doi.org/10.1098/rstb.2011.0244

Korhonen, L., Hadi, Packalen, P., & Rautiainen, M. (2017). Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index. *Remote Sensing of Environmen*t, 195, 259-274. https://doi.org/10.1016/j.rse.2017.03.021

Kross, A., McNairn, H., Lapen, D., Sunohara, M., & Champagne, C. (2015). Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops. *International Journal of Applied Earth Observation and Geoinformation*, 34(1), 235-248. https://doi.org/10.1016/j.jag.2014.08.002

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

Kumar, R., & Silva, L. (1973). Light Ray Tracing Through a Leaf Cross Section. *Applied Optics*, 12(12), 2950-2954. https://doi.org/10.1364/ao.12.002950

Lemus, R., & Lal, R. (2005). Bioenergy crops and carbon sequestration. In *Critical Reviews in Plant Sciences,* 24(1), 1-21. https://doi.org/10.1080/07352680590910393

Li, F., Zeng, Y., Luo, J., Ma, R., & Wu, B. (2016). Modeling grassland aboveground biomass using a pure vegetation index. *Ecological Indicators*, 62, 279-288. https://doi.org/10.1016/j.ecolind.2015.11.005

Lichtenthaler, H. K. (1987). Chlorophylls and Carotenoids: Pigments of Photosynthetic Biomembranes. *Methods in Enzymology*, 148, 350-382. https://doi.org/10.1016/0076-6879(87)48036-1

Litton, C. M., Sandquist, D. R., & Cordell, S. (2006). Effects of non-native grass invasion on aboveground carbon pools and tree population structure in a tropical dry forest of Hawaii. *Forest Ecology and Management*, 231(1-3), 105-113. https://doi.org/10.1016/j.foreco.2006.05.008

Lu, D. (2006). The potential and challenge of remote sensing-based biomass estimation. In *International*

*Journal of Remote Sensing*, 27(7), 1297-1328. https://doi.org/10.1080/01431160500486732

Marshall, M., & Thenkabail, P. (2015). Advantage of hyperspectral EO-1 Hyperion over multispectral IKONOS, GeoEye-1, WorldView-2, Landsat ETM+, and MODIS vegetation indices in crop biomass estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108, 205-218. https://doi.org/10.1016/j.isprsjprs.2015.08.001

Mascaro, J., Litton, C. M., Hughes, R. F., Uowolo, A., & Schnitzer, S. A. (2011). Minimizing Bias in Biomass Allometry: Model Selection and Log-Transformation of Data. *Biotropica*, 43(6). https://doi.org/10.1111/j.1744-7429.2011.00798.x

Mathew, I., Shimelis, H., Mutema, M., & Chaplot, V. (2017). What crop type for atmospheric carbon sequestration: Results from a global data analysis. *Agriculture, Ecosystems and Environment*, 243, 34-46. https://doi.org/10.1016/j.agee.2017.04.008

Merzlyak, M. N., Solovchenko, A. E., & Gitelson, A. A. (2003). Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in apple fruit. *Postharvest Biology and Technology*, 27(2), 197-211. https://doi.org/10.1016/S0925-5214(02)00066-2

Mutanga, O., & Skidmore, A. K. (2004). Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing*, 25(19), 3999-4014. https://doi.org/10.1080/01431160310001654923

National Aeronautics and Space Administration (NASA) (2013) The Electromagnetic Spectrum. Retrieved 1 April, 2020, from NASA Science website: https://imagine.gsfc.nasa.gov/science/toolbox/spectra1.html

Niechayev, N. A., Jones, A. M., Rosenthal, D. M., & Davis, S. C. (2019). A model of environmental limitations on production of Agave americana L. grown as a biofuel crop in semi-arid regions. *Journal of Experimental Botany*, 70(22), 6549-6559. https://doi.org/10.1093/jxb/ery383

Niklas, K. J. (2004). Plant allometry: Is there a grand unifying theory? In *Biological Reviews of the Cambridge Philosophical Society*, 79(4). https://doi.org/10.1017/S1464793104006499

Panda, S. S., Ames, D. P., & Panigrahi, S. (2010). Application of vegetation indices for agricultural crop yield prediction using neural network techniques. *Remote Sensing*, 2(3), 673-696. https://doi.org/10.3390/rs2030673

Pandit, S., Tsuyuki, S., & Dube, T. (2018). Estimating above-ground biomass in sub-tropical buffer zone community forests, Nepal, using Sentinel 2 data. *Remote Sensing*, 10(4), 601. https://doi.org/10.3390/rs10040601

Paul, K. I., Roxburgh, S. H., Chave, J., England, J. R., Zerihun, A., Specht, A., Lewis, T., Bennett, L. T., Baker, T. G., Adams, M. A., Huxtable, D., Montagu, K. D., Falster, D. S., Feller, M., Sochacki, S., Ritson, P., Bastin, G., Bartle, J., Wildy, D., Hobbs, T., Larmour, J., Waterworth, R., Stewart, H. T. L., Jonson, J., Forrester, D. I., Applegate, G., Mendham, D., Bradford, M., O'Grady, A., Green, D., Sudmeyer, R., Rance, S. J., Turner, J., Barton, C., Wenk, E. H., Grove, T., Attiwill, P. M., Pinkard, E., Butler, D., Brooksbank, K., Spencer, B., Snowdon, P., O'Brien, N., Battaglia, M., Cameron, D. M., Hamilton, S., McAuthur, G., & Sinclair, J. (2016). Testing the generality of above-ground biomass allometry across plant functional types at the continent scale. *Global Change Biology*, *22*(6), 2106–2124. https://doi.org/doi:10.1111/gcb.13201

Peichl, M., & Arain, M. A. (2007). Allometry and partitioning of above- and belowground tree biomass in an age-sequence of white pine forests. *Forest Ecology and Management*, 253(1-3), 68-80. https://doi.org/10.1016/j.foreco.2007.07.003

Pellikka, P. K.E., Heikinheimo, V., Hietanen, J., Schäfer, E., Siljander, M., & Heiskanen, J. (2018). Impact of land cover change on aboveground carbon stocks in Afromontane landscape in Kenya. *Applied Geography*, 94, 178-189. https://doi.org/10.1016/j.apgeog.2018.03.017

Pellikka, Petri K.E., Clark, B. J. F., Gosa, A. G., Himberg, N., Hurskainen, P., Maeda, E., Mwang'ombe, J., Omoro, L. M. A., & Siljander, M. (2013). Agricultural Expansion and Its Consequences in the Taita Hills, Kenya. In *Developments in Earth Surface Processes*, 16, 165-179. https://doi.org/10.1016/B978-0-444-59559-1.00013-X

Pérez-Pimienta, J. A., López-Ortega, M. G., & Sanchez, A. (2017). Recent developments in Agave performance as a drought-tolerant biofuel feedstock: agronomics, characterization, and biorefining. In *Biofuels, Bioproducts and Biorefining*, 11(4). https://doi.org/10.1002/bbb.1776

Picard, N., Saint-Andre, L., & Henry, M. (2012). Manual for building tree volume and biomass allometric equations: from field measurement to prediction. Available at: http://www.fao.org/3/i3058e/i3058e.pdf

Platts, P. J., Burgess, N. D., Gereau, R. E., Lovett, J. C., Marshall, A. R., McClean, C. J., Pellikka, P. K. E., Swetnam, R. D., & Marchant, R. (2011). Delimiting tropical mountain ecoregions for conservation. *Environmental Conservation*, 38(3), 312-324. https://doi.org/10.1017/S0376892911000191

Prabhakara, K., Hively, W. D., & McCarty, G. W. (2015). Evaluating the relationship between biomass, percent groundcover and remote sensing indices across six winter cover crop fields in Maryland, United States. *International Journal of Applied Earth Observation and Geoinformation*, *39*, 88–102. https://doi.org/10.1016/J.JAG.2015.03.002

Prudnikova, E., Savin, I., Vindeker, G., Grubina, P., Shishkonakova, E., & Sharychev, D. (2019). Influence of soil background on spectral reflectance of winter wheat crop canopy. *Remote Sensing*, 11(16), 1932. https://doi.org/10.3390/rs11161932

Python Software Foundation (2019). Python Language Reference, version 3.6. Available at http://www.python.org

Rasse, D. P., Rumpel, C., & Dignac, M. F. (2005). Is soil carbon mostly root carbon? Mechanisms for a specific stabilisation. *Plant and Soil*, 269, 341-356. https://doi.org/10.1007/s11104-004-0907-y

Raymond Hunt, E., Rock, B. N., & Nobel, P. S. (1987). Measurement of leaf relative water content by infrared reflectance. *Remote Sensing of Environment*, 22(3), 429-435. https://doi.org/10.1016/0034-4257(87)90094-0

Reddy, V. R., Pachepsky, Y. A., & Whisler, F. D. (1998). Allometric relationships in field-grown soybean. *Annals of Botany*, 82(1), 125-131. https://doi.org/10.1006/anbo.1998.0650

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Revill, A., Sus, O., Barrett, B., & Williams, M. (2013). Carbon cycling of European croplands: A framework for the assimilation of optical and microwave Earth observation data. *Remote Sensing of Environment*, 137, 84-93. https://doi.org/10.1016/j.rse.2013.06.002

Riihimäki, H., Luoto, M., & Heiskanen, J. (2019). Estimating fractional cover of tundra vegetation at multiple scales using unmanned aerial systems and optical satellite data. *Remote Sensing of Environment*, 224, 119-132. https://doi.org/10.1016/j.rse.2019.01.030

Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, 55(2), 95-107. https://doi.org/10.1016/0034-4257(95)00186-7

Rouse, W., Haas, R. H., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS, NASA SP-351. *Third ERTS-1 Symposium, Vol. 1*. Available at:

https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19740022614.pdf

Ruppert, D. (2004). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *Journal of the American Statistical Association*, 99(466), 567. https://doi.org/10.1198/jasa.2004.s339

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., Hagen, S., Petrova, S., White, L., Silman, M., & Morel, A. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences of the United States of America*, 108(24), 9899-9904. https://doi.org/10.1073/pnas.1019576108

Sahu, P., & Gupta, M. K. (2017). Sisal (Agave sisalana) fibre and its polymer-based composites: A review on current developments. *Journal of Reinforced Plastics and Composites*, *36*(24), 1759–1780. https://doi.org/10.1177/0731684417725584

Sampaio, E. V. S. B., & Silva, G. C. (2005). Biomass equations for Brazilian semiarid caatinga plants. *Acta Botanica Brasilica*, 19(4), 1677-941X.https://doi.org/10.1590/s0102-33062005000400028

Santos, J. D. G., Vieira, I. J. C., Braz-Filho, R., & Branco, A. (2015). Chemicals from Agave sisalana Biomass: Isolation and Identification. *International Journal of Molecular Sciences*, *16*(4), 8761–8771. https://doi.org/10.3390/ijms16048761

Sartori, F., Lal, R., Ebinger, M. H., & Parrish, D. J. (2006). Potential Soil Carbon Sequestration and CO2 Offset by Dedicated Energy Crops in the USA. *Critical Reviews in Plant Sciences*, *25*(5), 441–472. https://doi.org/10.1080/07352680600961021

Schlemmera, M., Gitelson, A., Schepersa, J., Fergusona, R., Peng, Y., Shanahana, J., & Rundquist, D. (2013). Remote estimation of nitrogen and chlorophyll contents in maize at leaf and canopy levels. *International Journal of Applied Earth Observation and Geoinformation*, 25, 47-54. https://doi.org/10.1016/j.jag.2013.04.003

*Sentinel-2 User Handbook* (2020). European Space Agency. Retrieved 1 April, 2020, from https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook

Serrano, L., Filella, L., & Penuelas, J. (2000). Remote Sensing of Biomass and Yield of Winter Wheat. *Crop Science*, 40(3). https://doi.org/10.2135/cropsci2000.403723x

Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples).

*Biometrika*, 52(3-4), 591-611. https://doi.org/10.2307/2333709

Sibanda, M., Mutanga, O., & Rouget, M. (2015). Examining the potential of Sentinel-2 MSI spectral resolution in quantifying above ground biomass across different fertilizer treatments. *ISPRS Journal of Photogrammetry and Remote Sensing*, 110, 55-56. https://doi.org/10.1016/j.isprsjprs.2015.10.005

Silleos, G., Alexandridis, T., Gitas, I., & Perakis, K. (2006). Vegetation indices: Advances made in biomass estimation and vegetation monitoring in the last 30 years. *Geocarto International*, 21(4), 21-28. https://doi.org/10.1080/10106040608542399

Singh, B. P. (2013). *Biofuel crops: production, physiology and genetics, Chapter 4, 292.* CAB eBooks. https://doi.org/10.1079/9781845938857.0000

Smith, B. W., & Brand, G. J. (1983). *Allometric biomass equations for 98 species of herbs, shrubs, and small trees*. USDA Forest Service.

Smith, P. (2006). Soils as carbon sinks: the global context. *Soil Use and Management*, 20(2). https://doi.org/10.1111/j.1475-2743.2004.tb00361.x

Terrapon-Pfaff, J. C., Fischedick, M., & Monheim, H. (2012). Energy potentials and sustainability—the case of sisal residues in Tanzania. *Energy for Sustainable Development*, 16(3), 312–319. https://doi.org/10.1016/J.ESD.2012.06.001

Tilly, N., Aasen, H., & Bareth, G. (2015). Fusion of plant height and vegetation indices for the estimation of barley biomass. *Remote Sensing*, 7(9), 11449-11480. https://doi.org/10.3390/rs70911449

Tucker, C. J. (1980). A critical review of remote sensing and other methods for non-destructive estimation of standing crop biomass. In *Grass and Forage Science*, 35(3). https://doi.org/10.1111/j.1365-2494.1980.tb01509.x

Viña, A., Gitelson, A. A., Nguy-Robertson, A. L., & Peng, Y. (2011). Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sensing of Environment*, 115(12), 3468-34-78. https://doi.org/10.1016/j.rse.2011.08.010

Von Cruz, M. V., & Dierig, D. A. (2015). *Industrial Crops: Breeding for Bioenergy and Bioproducts*. https://doi.org/10.1007/978-1-4939-1447-0

Wang, L., Zhou, X., Zhu, X., Dong, Z., & Guo, W. (2016). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *Crop Journal*, 4(3), 212–219. https://doi.org/10.1016/j.cj.2016.01.008

Warton, D. I., Wright, I. J., Falster, D. S., & Westoby, M. (2006). Bivariate line-fitting methods for allometry. *Biological Reviews*, 81(2), 259–291. https://doi.org/doi:10.1017/S1464793106007007

Wickham, H. (2009) *ggplot2: elegant graphics for data analysis*. Springer New York. Version 3.2.1

Wiseman, G., McNairn, H., Homayouni, S., & Shang, J. (2014). RADARSAT-2 Polarimetric SAR response to crop biomass for agricultural production monitoring. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(4), 4461-4471. https://doi.org/10.1109/JSTARS.2014.2322311

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R, Second Edition*. New York. https://doi.org/10.1201/9781315370279

Wood, S. N. (2017). *mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation*. Version 1.8-31

Youkhana, A. H., Ogoshi, R. M., Kiniry, J. R., Meki, M. N., Nakahata, M. H., & Crow, S. E. (2017). Allometric Models for Predicting Aboveground Biomass and Carbon Stock of Tropical Perennial C4 Grasses in Hawaii . In *Frontiers in Plant Scien*e, 8, 650. https://www.frontiersin.org/article/10.3389/fpls.2017.00650

Zan, C. S., Fyles, J. W., Girouard, P., & Samson, R. A. (2001). Carbon sequestration in perennial bioenergy, annual corn and uncultivated systems in southern Quebec. *Agriculture, Ecosystems and Environment*, 84, 135-144. https://doi.org/10.1016/S0167-8809(00)00273-5

# APPENDICES

**Appendix 1**. Details of the linear regressions. (A) Model 1, (B) Model 2.

**A.**

```
Call:
lm(formula = log(dryw_t) ~ log(diam_max^2 * length), data = sisal)

Residuals:
     Min       1Q   Median       3Q      Max
-0.28883 -0.12787 -0.03344  0.13055  0.28070

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)              -5.80251    0.21477  -27.02   <2e-16 ***
log(diam_max^2 * length)  0.92666    0.02106   44.00   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.17 on 36 degrees of freedom
Multiple R-squared:  0.9817,    Adjusted R-squared:  0.9812
F-statistic:  1936 on 1 and 36 DF,  p-value: < 2.2e-16
```

**B.**

```
Call:
lm(formula = log(dryw_t) ~ log(diam_max^2 * plant_h), data = sisal)

Residuals:
      Min        1Q    Median        3Q       Max
-0.297419 -0.085384  0.001117  0.089558  0.198489

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               -5.42299    0.14701  -36.89   <2e-16 ***
log(diam_max^2 * plant_h)  0.84999    0.01377   61.73   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1217 on 36 degrees of freedom
Multiple R-squared:  0.9906,    Adjusted R-squared:  0.9904
F-statistic:  3810 on 1 and 36 DF,  p-value: < 2.2e-16
```

**Appendix 2.** Summary of the RE2/RE3 GAM.

```
Formula:
biomass ~ s(RE2_RE3_rsi, k = 3)

Parametric coefficients:
            Estimate Std. Error t value            Pr(>|t|)
(Intercept)  12.2446     0.6165   19.86 <0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                 edf Ref.df     F             p-value
s(RE2_RE3_rsi) 1.726  1.925 84.44 <0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =   0.73   Deviance explained = 73.8%
GCV = 23.134  Scale est. = 22.047    n = 58
```