Student thesis series INES nr 522

Modelling gross primary production in semi-arid regions: Effects on carbon uptake of intensive agriculture in southern Kenya

Malin Ahlbäck

2020

Department of

Physical Geography and Ecosystem Science

Lund University

Sölvegatan 12

S-223 62 Lund

Sweden

Malin Ahlbäck (2020).
Title of thesis (English)
Modelling gross primary production in semi-arid regions: Effects on carbon uptake of intensive agriculture in southern Kenya
Title of thesis in Swedish
Modellera bruttoprimärproduktion i semi-arida miljöer: Effekter på kolupptag från intensivt jordbruk i södra Kenya
Master degree thesis, 30 credits in *Physical Geography and Ecosystem Science*
Department of Physical Geography and Ecosystem Science, Lund University

Level: Master of Science (MSc)

Course duration: *January* 2020 until *June* 2020

Disclaimer

Modelling gross primary production in semi-arid regions: Effects on carbon uptake of intensive agriculture in southern Kenya

---

Malin Ahlbäck

Master thesis, 30 credits, in *Physical Geography and Ecosystem Science*

Supervisor Lars Eklundh
Department of Physical Geography and Ecosystem Science,
Lund University

Exam committee:
Jonas Ardö,
Department of Physical Geography and Ecosystem Science,
Lund University

Virginia Garcia-Millan,
Department of Physical Geography and Ecosystem Science,
Lund University

Acknowledgements

Many people have in one way or another contributed to this work. First of all, I would like to thank my supervisor Lars Eklundh for introducing me to this topic, and for the very helpful and insightful advices you have given along the way. Second, I would like to thank my examiners Jonas Ardö and Virginia Garcia-Millan for their valuable feedback. I would also like to thank Janne Rinne, firstly for supplying the flux tower data from the Ausquest farm, and secondly for helping me with the pre-processing of the flux data. Many thanks also to Zhanzhang Cai and Torbern Tagesson, for taking the time to help me with Sentinel-2 processing in TIMESAT and to answer questions regarding the GPP model.

Writing a thesis is a challenge, but I would argue that writing a thesis during a global pandemic is taking it to a new level. To all my wonderful classmates who have supported and encouraged me during this semester and long before that: Thank you! I have had a blast with you for the past 2 years. I wish you all the best on your future endeavours.

Abstract

Background and aim: Gross primary production (GPP) is the largest global carbon (C) flux and an important component for counteracting anthropogenic CO_2 emissions, understanding vegetation dynamics, and sustaining universal human standards. Africa plays a prominent role in the global C cycle, though our understanding of GPP dynamics is largely hampered by a paucity of ground-based observations. By using flux data from a newly installed eddy covariance (EC) flux tower in southern Kenya and high-resolution Sentinel-2 satellite data, the aims of this thesis were to i) determine the net C flux and climatic factors governing GPP at a high-intensively cropped dryland farm, ii) use the dataset to test different remote sensing driven GPP models and the MOD17 GPP product (collection 6), and iii) assess the difference in annual GPP among croplands and shrublands, grasslands and open forests.

Results: The studied cropland had a net ecosystem exchange (NEE) of 170 g C m^{-2} between March and December 2019, thus acting as a source of C during the measurement period. Climatic factors governing GPP were vapour pressure deficit (VPD), soil moisture and air temperature (T_{air}). Using the 2-band enhanced vegetation index (EVI2), VPD and T_{air} as input in a multiple linear regression yielded the best results to in situ GPP (R^2=0.65, RMSE = 2.02 g C m^{-2} d^{-1}). Croplands had a higher mean annual GPP compared to grasslands, shrublands and open forest in the sub-humid to semi-arid agro-climatic zone, and a similar annual GPP to open forest in the semi-arid agro-climatic zones. The studied field at the Ausquest farm had however a considerably lower GPP compared to both natural lands and other croplands.

Conclusion: Based on these results, land cover changes occurring over the vast savannah landscape in southern Kenya are not likely to reduce GPP in the region, however, more data is needed to make a more robust assessment. The studied cropland at the Ausquest farm emitted considerable amounts of C during the measurement period, indicating the importance of accounting for respiration in land-cover change studies. The ambiguity of the GPP models can be attributed to the persistent cloud cover during the growing season, as well as difficulties in accounting for the influence of anthropogenic effects on GPP, and to little ground-based data for calibration and validation.

***Keywords**: Physical Geography; Ecosystem Analysis; GPP; Africa; Kenya; Land cover changes; Cropland; Remote sensing; Sentinel-2*

Sammanfattning

Bakgrund och syfte: Bruttoprimärproduktion (GPP) är det största globala kol (C)-flödet och en viktig komponent för att motverka antropogena CO_2-utsläpp, förstå vegetationsdynamik och för att upprätthålla universella mänskliga levnadsstandarder. Afrika har en framträdande roll i den globala C-cykeln, även om vår förståelse av dess GPP-dynamik till stor del hindras av en bristande tillgång på markbaserade observationer. Genom att använda data från ett nyinstallerat eddy kovarians (EC)-torn i södra Kenya och Sentinel-2-data med hög rumslig upplösning, syftar denna uppsats på att i) bestämma C-balansen på ett högintensivt jordbruksfält samt vilka klimatfaktorer som påverkar GPP, ii) testa och jämföra olika fjärranalys-baserade GPP-modeller, och iii) göra en bedömning av skillnaden i GPP mellan grödor och de naturliga, semi-arida ekosystemen.

Resultat: Den studerade jordbruksmarken hade ett nettoekosystemutbyte (NEE) på 170 g C m^{-2} mellan mars-december 2019, vilket indikerar att det var en källa till C under mätperioden. Klimatfaktorer som styrde GPP var ångtryckunderskott (VPD), markfuktighet och lufttemperatur. Användning av det 2-bandsförbättrade vegetationsindexet (EVI2), VPD och lufttemperatur i en multipel linjär regression gav det bästa resultaten till in situ GPP (R^2 = 0,65, RMSE = 2,02 g C m^{-2} d^{-1}). Genom att skala upp modellen över södra Kenya jämfördes skillnaden i årlig GPP mellan jordbruksmark, busklandskap, gräsmark och öppen skog. Jordbruksmarker hade generellt en högre genomsnittlig årlig GPP jämfört med busklandskap, gräsmark och öppen skog i den semi-humida till semi-arida klimatzonerna. I den semi-arid zonen var dock årligt GPP mellan jordbruksmark och öppen skog lika. Årlig GPP var korrelerad med antalet dagar med GPP> 2 g C m^{-2} d^{-1}, men jordbruksmark och gräsmark hade en något lägre korrelation, vilket indikerar att växtsäsongens längd till stor del påverkade den årliga GPP.

Slutsats: Baserat på dessa resultat skulle en ökning av jordbruksmark på det stora savannlandskapet i södra Kenya sannolikt inte minska GPP, men mer information behövs för att göra en robust bedömning. Den studerade jordbruksmarken släppte ut en avsevärd mängd C under mätperioden, vilket indikerar vikten av respiration i markförändringsstudier. Tvetydigheten i GPP-modellerna kan hänföras till molnighet under växtsäsongen, liksom svårigheterna att fånga förändringar i GPP härledd till antropogen påverkan, samt den limiterade tillgången på in situ-data för kalibrering och validering.

Contents

List of abbreviations

ACZ	Agri-Climatic Zone
CO_2	Carbon Dioxide
EC	Eddy Covariance
EO	Earth Observations
ESA	European Space Agency
EVI	Enhanced Vegetation Index
EVI2	Enhanced 2-band Vegetation Index
FAO	Food and Agriculture Organisation of the United Nations
GPP	Gross Primary Production
IAV	Inter-Annual Variability
LC	Land Cover
LUE	Light-Use Efficiency
MAP	Mean Annual Precipitation
MLR	Multiple Linear Regression
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetation Index
NEE	Net Ecosystem Exchange
NL	Non-linear asymptotic Mitscherlich ligh-response curve
NPP	Net Primary Production
OLS	Ordinary Least Squares
PAR	Photosynthetically Active Radiation
R_{eco}	Ecosystem Respiration
RMSE	Root Mean Squared Error
RS	Remote Sensing
S2	Sentinel-2
SM	Soil Moisture
VPD	Vapour Pressure Deficit

1. Introduction

Gross primary production (GPP), the carbon (C) uptake by ecosystems through the process of photosynthesis, is the largest global C flux and driver of ecosystem functions such as plant growth and respiration (Beer et al. 2010). GPP, together with respiration, regulates land-atmosphere CO_2 exchange, consequently being an important component for counteracting human-induced C emissions. Furthermore, capturing variations in GPP in both space and time becomes of high societal relevance against the backdrop of anthropogenic influences, such as land-cover (LC) and climate change, ultimately altering the terrestrial C balance (Beer et al. 2010; Houghton et al. 2012).

Since the release of satellites for Earth observations (EO) in the 1970s, remote sensing (RS) has become a key method to monitor our planet's land surfaces at increasingly finer spatiotemporal resolutions. Today, several RS driven GPP models exist and sustain an increasing understanding of terrestrial vegetation dynamics. However, many of these models use data at a resolution of hundred metres or more, making it difficult to match the sample support size from satellite products to in situ GPP derived from eddy covariance (EC) stations (Tan et al. 2006), especially for very heterogeneous ecosystems such as savannahs. With the launch of the Sentinel-2 (S2) satellite constellation in 2015, new possibilities to delineate ecosystem structures at a spatial resolution up to 10 metres are within reach.

African drylands account for a large proportion of the global inter-annual variability in C levels and play a prominent role in the global terrestrial C balance (Poulter et al. 2014; Tagesson et al. 2020; Ahlström et al. 2015). However, large deviations in the GPP estimation for Africa, and hence the globe, persist. As stated by Williams et al. (2007), it is "*one of the weakest links in our understanding of the global C cycle*". This is partly attributed to the relatively little attention Africa's semi-arid ecosystems have been subject to in C studies, and the scarcity of ground-based data, such as EC and climate stations (Ciais et al. 2011; Ardö et al. 2008). Altogether, these factors hamper the development of GPP models.

The research site at Ausquest farm was established in March 2019 as one of few sites for measuring C dynamics using the EC technique in Kenya, Eastern Africa. As such, this dataset boosts the so-far lacking C monitoring network over the region by creating opportunities to improve the knowledge on semi-arid cropland ecosystems, and to use the data for RS upscaling purposes. Situated in a region undergoing LC changes, with fast appropriation of the savannah landscape, it is also of interest to compare the cropland's C cycle to natural drylands.

Using this first-hand dataset of continuous C flux measurements, the aims of this thesis are to determine the net C flux and climatic factors governing GPP at the Ausquest farm, and to use this dataset to test different RS driven GPP models for upscaling purposes. By upscaling in situ GPP using high-resolution S2 satellite data, this thesis presents an assessment on the difference in GPP between croplands and natural semi-arid drylands in southern Kenya. A more detailed description of the aims is given at the end of the succeeding chapter (2.4).

2. State of science

This chapter introduces concepts and research that have led to the aims of this thesis. The body of literature encountered is largely confined to either small-scale local studies, or broad-scale RS studies over the Sub-Saharan region or the Sahel. As such, though this thesis aims to focus on the southern region of Kenya, the background is largely inferred from the Sahelian belt, situated between the Sahel desert and the sub-humid equatorial belt, or Sub-Sahara, encompassing the region south of the Sahel.

2.1 In situ GPP estimates

GPP cannot be directly measured. However, with ground-based, high-frequency measurements of the vertical CO_2 flux density, derived from the covariance between the vertical wind component and the measured gas concentration, the EC method captures the net ecosystem exchange (NEE), from which the GPP can be estimated:

$$NEE = GPP + R_{eco} \qquad Eq.\ 1$$

where R_{eco} is the ecosystem respiration, i.e. the sum of heterotrophic and autotrophic respiration. From the NEE, it can thus also be determined whether an ecosystem is acting as a sink (negative NEE) or source (positive NEE) of C on various time scales.

The EC technique plays a prominent role in global C cycling studies. Several national and international EC networks exist (e.g. ICOS, FluxNet, CarboAfrica, EuroFlux and AmeriFlux), providing in situ C flux measurements for calibration and validation of RS-driven GPP models. However, though covering a large fraction of the Earth's surface, the flux-site density and number is lower over Africa compared to other continents, hampering the development of our understanding of the system in semi-arid regions (Ciais et al. 2011; Ardö et al. 2008).

2.2 Modelling GPP in semi-arid regions from space

By combining EO data with mathematical functions, uptake and/or release of C from land can be estimated through space and time, allowing for assessment of changes and quantities in the terrestrial C pool on various scales. This section will describe important constrains on GPP in semi-arid regions that are incorporated in RS-driven GPP models, different concepts and methods to model GPP, and their respective advantages and limitations.

2.2.1 Environmental drivers and constrains on GPP

Water availability, or rather deficit, is the strongest driver of inter-annual variability over the African continent and unarguable a major factor governing variations in GPP in semi-arid

environments (Hickler et al. 2005). Variables describing water availability is therefore often incorporated in RS driven GPP models, though the underlying physiological relationships regulating assimilation on a leaf-to-canopy level remain to be disintegrated. The vapour pressure deficit (VPD), describing the difference between the water vapour pressure at saturation and the actual vapor pressure of the air at a given temperature, is an important driver of stomatal conductance (Oren et al. 1999). Increased VPD in the air causes the stomata to close in order to prevent water losses, consequently reducing the plant's efficiency to take up CO_2. Studies over the Sahel have shown that this limit, from where GPP almost ceases, is approximately at a VPD above 20 hPa (Merbold et al. 2009; Abdi et al. 2017). However, both the shallowest and the deepest root systems occur in semi-arid ecosystems (Fan et al. 2005) and access to deep water by roots has shown to enable assimilation of C even during dry periods of some species (Ardö et al. 2008; Leuning et al. 2005). Today's remote-sensing driven GPP models often incorporate VPD (e.g. Running et al. 2004; Kanniah et al. 2013) or EO data (Fensholt et al. 2006; Abdi et al. 2017) as a proxy for describing dryness. They do however seldom include the effects of soil moisture, especially not from deeper levels, which results in overestimations of GPP in drought conditions (Stocker et al. 2019).

2.2.2 The MODIS GPP product

The light-use-efficiency (LUE) concept (Monteith 1972) is based on the theory that GPP is the product of the absorbed photosynthetically active radiation (APAR) by the canopy layer, and the efficiency with which it converts the radiative energy into C, conveniently conforming the complexity of photosynthesis to a linear model:

$$GPP = PAR * fAPAR * LUE_{max} * LUE_s \qquad Eq.\ 2$$

where fAPAR is the fraction of the photosynthetically active radiation (PAR) absorbed by the canopy, LUE_{max} is the biological maximum LUE of the canopy layer, and LUE_s represent one or several environmental stress scalars that downregulates LUE in sub-optimal conditions. The fAPAR is typically derived using a vegetation index, while the LUE can be estimated indirectly from environmental stresses such as water or nutrient deficit, oxidative stress or low/high temperatures.

Based on the LUE concept, using VPD and air temperature as LUE_s, the currently only available near-real-time global dataset of GPP is the Moderate Resolution Imaging Spectroradiometer (MODIS) GPP product (MOD17). Several studies have evaluated the MOD17 GPP product. The dataset has fair agreement over boreal regions (Turner et al. 2005) but yields ambiguous estimates over drought-induced ecosystems. Ma et al. (2014) and Sjöström et al. (2011) found the Enhanced Vegetation Index (EVI) alone to be a better descriptor of GPP variability than the MOD17 GPP product over an Australian and Sahelian savannah, respectively. Studies have indicated that the biome-specific values determining the LUE in the model need adjustment (Sjöström et al. 2013) and that the LUE_{max} value is too low for savannah ecosystems (Tagesson et al. 2017). Some authors have also highlighted the lack of a soil water limiting factor on LUE as a reason for the uncertainty (Leuning et al. 2005; Stocker et al. 2019; Kanniah et al. 2009)

and the too coarse meteorological data (Fensholt et al. 2006). Due to this ambiguity, studies have developed alternative methods to model GPP in semi-arid regions.

2.2.3 Modelling GPP using a non-linear light-response curve

The LUE concept (Eq. 2) relies on the three assumptions that i) GPP is linearly related to APAR, ii) there is a direct link between a vegetation index and APAR, and iii) there is a biophysical reason behind the reduction in C uptake which can be explained by LUE_S. However, the biophysical relationship behind C assimilation with increased exposure to PAR is non-linear in nature on a leaf scale (Figure 1). From the light compensation point (the point on the light-response curve where the rate of photosynthesis matches the rate of cellular respiration) to approximately 200 µmol (photon) m^{-2} s^{-1}, there is a linear portion referred to as the maximum quantum yield (α), whereas beyond this range progressive curvature follows until photosynthesis reaches a semi-plateau, referred to as the maximum photosynthetical capacity at light saturation (Fopt) (Hall et al. 1994). Most RS studies have found, or assumed, a linear relationship between the APAR and GPP on a canopy scale. However, Tagesson et al. (2017) successfully applied a non-linear light-response GPP model by upscaling Fopt and α with vegetation indices derived from MODIS Terra over the Sahel, resulting in a higher performance compared to the MOD17 GPP product.

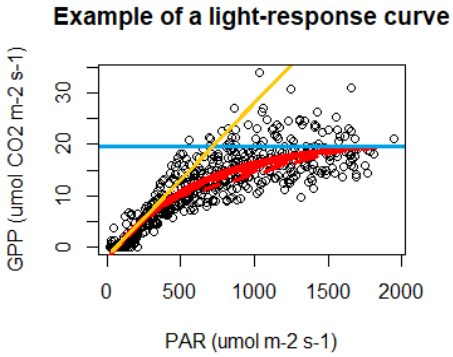

Figure 1: A typical light-response curve (red curve), showing the initial linear relationship between GPP and PAR called the maximum quantum yield (α; yellow line) and the gradual saturation reached when increasing PAR does not increase C assimilation, referred to as the maximum photosynthetical capacity (Fopt; blue line).

2.2.4 GPP models based on regression analysis

A common method to model GPP is to use regression analysis between EO data and in situ GPP. Abdi et al. (2017) displayed that a multiple linear regression (MLR) GPP model combining EVI, land surface temperature, and band 7 of the MODIS sensor explained 89% of the variability in GPP at three Sahelian EC sites. Though rather applied as a way to compare the effect of exchanging climate data in the MOD17 algorithm with the shortwave infrared water stress index (SIWSI), Fensholt et al. (2006) showed a MLR model to perform better than the MOD17 product over a Sahelian site. Other studies have used MLR for modelling GPP in Northern forests (Schubert et al. 2012) and phenology of savannahs (Boke-Olén et al. 2016) with good results. Being an empirical model, it has limitations in terms of upscaling the calibrated fit to other locations and times (Hilker et al. 2008). However, other advantages are that it is simple, and has a methodology that allows for more freely scaling and incorporation of environmental constraints on GPP compared to the traditional LUE method.

2.3 African semi-arid ecosystems and LC changes

About half of Africa is covered by savannahs (Allaby 2006). Characteristic for these landscapes is their continuous tropical grassland vegetation, with an open discontinuous canopy consisting of woody shrubs or trees. The woodland-grassland dynamic makes savannahs highly transient systems (Bond et al. 2005), whose plant structure and spatial extent is closely yet complicatedly related to the mean annual precipitation (MAP), fire regime, grazing pressure by cattle and wildlife and anthropogenic influences (Sankaran et al. 2005; Bucini and Hanan 2007). Studies have found MAP to have a strong control on the percentage of woody vegetation (Sankaran et al. 2005), species richness (Tagesson et al. 2015), and maximum C assimilation rates (Merbold et al. 2009). As such, it is difficult to fully define the biome and disentangle all its broad variability into discrete classes; in this thesis, savannah landscape mosaic will be used when referring to a mixture of plant structures or ecosystems typical of savannahs, such as shrubs, open tree canopies, and continuous grass cover.

Currently, little research is available on the consequences of LC changes on the C balance in semi-arid Africa, partly since the C dynamics of the systems being appropriated are rather poorly understood to begin with (Ciais et al. 2011). RS studies on LC changes have shown an annual increase in agricultural land of 1.4% in East Africa (years 1990-2010), of which 1.2% are derived from savannahs and grasslands (Brink et al. 2014), marking the importance of understanding consequences of these changes. The MOD17 GPP and NPP (net primary production; the part of GPP allocated to plant growth) products and the dynamic vegetation model LPJ-GUESS indicate that savannahs have higher GPP and NPP compared to croplands on a continental scale, while shrublands and grasslands have lower ones (Ardö 2015). Others show that NPP of croplands is also larger than that of savannahs in Africa (Williams et al. 2008). Though only compared to one EC station situated on a cropland, Tagesson et al. (2016) found the GPP of a Millet farm to be lower compared to savannahs across the Sahel. Looking at trends in leaf area index (LAI, the unit leaf area per unit area (m^2 m^{-2})), Zhu et al. (2016) shows that LC changes have resulted in both positive and negative trends in LAI in semi-arid Kenya and Sub-Sahara at large. Thus, there seems to be little consensus on the difference, and hence impact, of cropland GPP versus savannah GPP.

2.4 Research objectives

Based on the research summarised in this chapter, the following focal aims have been derived:

1. To determine the net C flux and the climatic factors regulating GPP on a high-intensively cropped dryland farm, exemplified by the cropland at the Ausquest farm.

2. To compare and evaluate the performance of the existing GPP product MOD17A2 GPP (collection 6) and two different self-developed methods to model and upscale GPP: non-linear modelling based on the light-response curve, and multiple linear regression analysis.

3. To determine whether the annual GPP in 2019 differs between croplands and natural semi-arid land covers typical of the savannah landscape (grassland, shrubs and open forest) within the study area.

The first aim is fulfilled by statistically analysing in situ observations of C fluxes, VPD, air temperature and PAR from the Ausquest flux tower and modelled soil moisture data, spanning from March 2019 to December 2019.

The second aim is fulfilled following the methodology by Tagesson et al. (2017) to derive a non-linear model based on an asymptotic Mitscherlich light-response curve. Multiple linear regression models are calibrated using in situ GPP and modelled climate data. The models are thereafter compared to in situ GPP from the Ausquest farm using ordinary least square regression, and the best-performing model is selected for upscaling to achieve aim iii).

The third aim is fulfilled by comparing the annual GPP of different LC classes within the semi-humid to semi-arid and semi-arid zones using zonal statistics and visual interpretations.

3. Study area

The study area is confined to the extent of the imagery of the S2 tile 37MBU corresponding to the location of the EC station at Ausquest farm (Figure 2). The total area is approximately 12,000 km^2 and situated between 0°54′S and 1°53′S and 36°18′ E and 37°17′E in the southeastern part of Kenya. In the centre of the study area is the capital Nairobi, surrounded by western Machakos, southern Muranga and Nakuru, Kiambu, and north-eastern Kajiado.

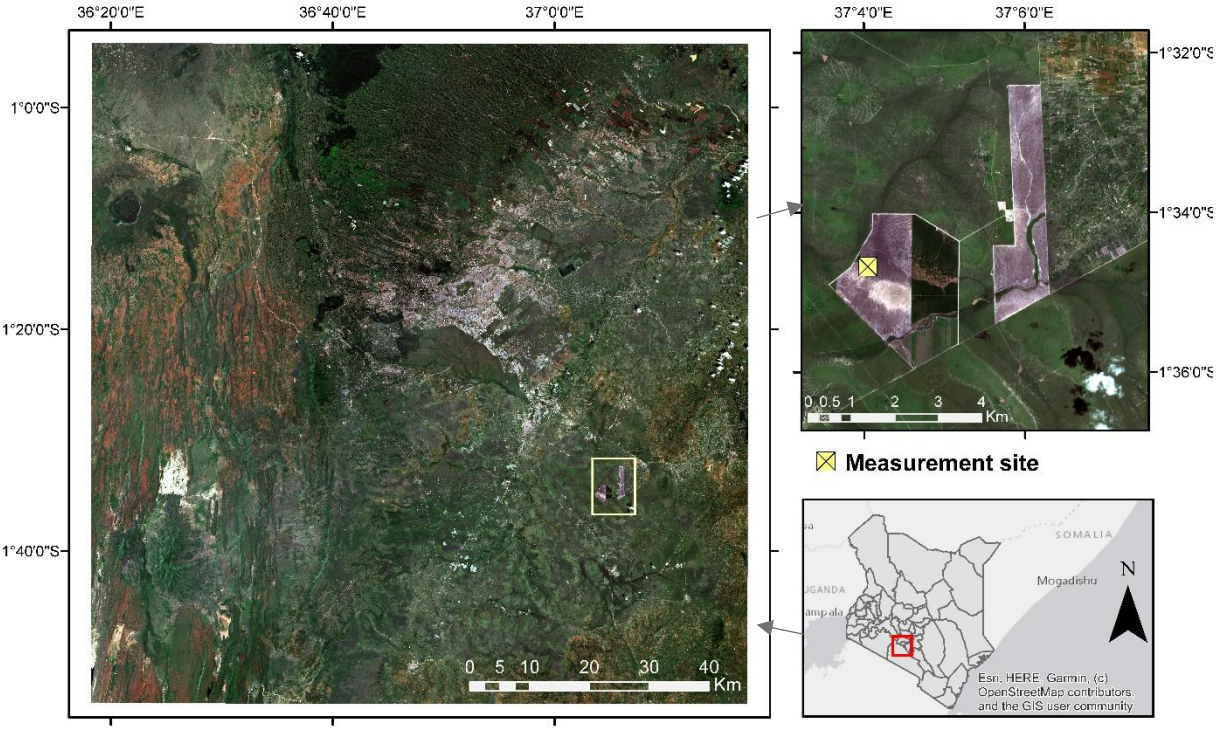

Figure 2: The full study area (left) defined by the Sentinel-2 tile. The EC measurement site at Ausquest farm is denoted with a yellow square in the tile, and as a yellow crossed square in the smaller image to the right. The location of the study area (tile) in Kenya is seen in the lower right image.

The climate is largely influenced by the inter-tropical convergence zone (ITCZ). Droughts or semi-drought events are an annually reoccurring event, followed by rainfall from March to May (long rain season) and October to December (short rain season). However, the nature of rainfall in the region is highly irregular in both space and time (Giannini et al. 2008). The mean annual temperature ranges between 16-30 ºC, depending on altitude (Sombroek et al. 1982).

The Kenyan highlands are situated north of Nairobi, and are sub-humid, high-productivity zones, therefore encompassing most of the cultivated land in the study area. The remaining region is sub-arid to arid, made up by a mosaic of savannah, grassland, woodland and farmlands.

The flux tower is situated approximately 40 km southeast from Nairobi in vicinity of the Athi river. The studied cropland is like the surrounding fields rain-fed and has minor fertilization, therefore roughly exposed to the same environmental constrains as the surrounding natural

vegetation. The site is underlain by black cotton soils, a clay rich soil, exerting strong expansion and shrinking characteristics when wet and dry. The cracks created during the dry season allow for organic material to pass down to deeper soil layers, creating a natural turnover for organic material (Bronswijk 1991). The Ausquest farm aims to maintain the sequestration of soil C and therefore utilises no-till technology, crop rotation and leaves a fraction of the biomass on the field after harvesting (Sigei 2020).

4. Data & methods

This study utilises a combination of in situ data, modelled climate data and S2 imageries to model GPP, as well as discrete spatial data for stratifying GPP into LC classes and climatic zones, all summarised in section 4.1. The workflow of the methods presented in section 4.2 is ordered in accordance to each of the three aims of this thesis, and can be generalised into: i) Data pre-processing, ii) Determining the importance of environmental factors, iii) Model development and evaluation, and iv) Upscaling GPP, LC stratification and statistical analysis. The data analyses were carried out in R 3.6.2 (R Development Core Team, 2020), unless otherwise stated.

4.1 Data

4.1.1 Sentinel-2 data

The S2 mission consists of a two-satellite constellation, which together enable a revisiting time of 5 days close to the equator, and operates at a swath width of 290 km. The multispectral instrument onboard the satellites samples in 13 bands from 443-2190 nm with a spatial resolution up to 10 metres, thus outperforming many other satellite products in terms of spatial resolution.

The S2 level 2A product of tile 37MBU was downloaded from the official data hub provided by ESA (www.scihub.copernicus.eu/dhus). In total, 72 images were available between December 2018 to January 2020. The level 2A product contains geometrically corrected images of bottom of atmosphere (i.e. surface) spectral reflectance, hence lowering the need for further processing. Yet, noise caused from cloud cover and other disturbances is still present, and therefore time series smoothing was implemented prior to the analysis (see section 4.3.1).

The Enhanced Vegetation Index (EVI) has been found to outperform other indices for modelling GPP over semi-arid regions (Sjöström et al. 2009; Ma et al. 2014). However, with the advent of new methods to effectively correct for atmospheric disturbances, the inclusion of a blue band has been argued to rather introduce more uncertainty as the scattering in the blue wavelengths is high for dense green vegetated areas (Yengoh et al. 2015). Thus, the developed 2-band EVI (EVI2) (Jiang et al. 2008) was used in this study:

$$EVI2 = 2.4\ x \frac{NIR - RED}{NIR + 2.4\ x\ RED + 1} \quad \text{Eq. 3}$$

Where NIR refers to the near-infrared band (band 8) and the red to the red band (band 4) of the S2 product, both sampled at a spatial resolution of 10 m.

4.1.2 In situ measurements

Though not directly used as input to the models, in situ data was used to calibrate or validate the models and their respective spatial data inputs. Data from the flux tower (1°34'40.7"S, 37°4'3.0"E) and ground-based spectral tower (1°34'40.9"S, 37°04'02.9"E) was retrieved between the period March 2019 until the end of December 2019. Measurements of NEE (μmol CO_2 m^{-2} s^{-1}) and water vapor (mmol H_2O m^{-2} s^{-1}) were recorded at 20 Hz using an open-path CO_2/H_2O gas analyser (LI-7500, LI-COR Inc. Lincoln, Nebraska, USA) and a 3-D sonic Anemometer (GILL instruments UK) installed at a height of 3.2 m. A full summary of the set-up is found in Table 1.

Ancillary data of air temperature (T_{air}), relative humidity, and incoming PAR was measured at the flux site every 30 seconds and stored as 30 minutes means. Data of ground-based spectral reflectance was collected from a spectral sensor installed on a mast approximately 50 m from the flux tower site. Measurements of incoming and outgoing light were used to calculate ground based EVI2 as:

$$EVI2 = 2.5\ x \frac{\rho NIR - \rho RED}{\rho NIR + 2.4\ x\ \rho RED + 1} \quad \text{Eq. 4}$$

where ρNIR and ρRED are the hemispherical reflectance in the NIR (channel 2) and red (channel 1) bands, respectively.

Table 1: A summary of the set-up of the site-level measurements used in this study. PAR stands for photosynthetically active radiation.

Variable	Measurement height (m)	Sensor	Company
Air temperature (ºC)	2	HMP155	Viasala, Helsinki, Finland
NEE / H_2O (μmol CO_2 m^{-2} s^{-1}/ mmol H_2O m^{-2} s^{-1})	3.2	EC155	Campbell Scientific, Logan, UT, USA
PAR (μmol photons m^{-2} s^{-1})	3.5	LI-109/R	Li-Cor, Lincoln, Nebraska
Relative humidity (%)	2	HMP155	Viasala, Helsinki, Finland
Channel 1 (nm) (centre wavelength 656, band width 22)	8	SKR1860	Skye Instruments Ltd., Llandrindod Wells, UK
Channel 2 (nm) (centre wavelength 866, band width 29)	8	SKR1860	Skye Instruments Ltd., Llandrindod Wells, UK

4.1.3 Gridded climate data

Gridded climate was used as model input to upscale GPP over the region. Daytime values (07:00-19:00) between 1 January 2019 to 31 December 2019 of surface solar downwelling radiation (J m^{-2}), air temperature at 2 m height (K), dew point temperature at 2 m height (K), and volumetric soil moisture (m^3 m^{-3}) at four depths (0-7, 7-21, 21-72 and 72-189 cm) were retrieved from the European Centre for Medium-Range Weather Forecast (ECMWF) ERA5 reanalysis dataset. The ERA5 is the fifth and latest global climate reanalysis dataset produced by ECMWF, and a continuation of the ERA-Interim project. The data is produced by blending

a land-atmosphere-ocean model with observed meteorological data, resulting in a product with 0.1° x 0.1° spatial resolution and hourly time resolution.

All variables were aggregated into daily values to match the time step of the GPP model. To convert solar radiation to PAR (µmol m^{-2} s^{-1}), a factor of 0.45 was used to account for the proportion of the spectra present in PAR (McCree 1972). The daily mean of the total soil moisture (SM) (mm) of the soil column was calculated by taking the sum of the values from all four soil layers. The dew point temperature (T_d) and daily T_{air} were used to calculate VPD (hPa) following the formula by Allen et al. (1998):

$$VPD = e_s - e_a \quad \text{Eq. 5}$$

where e_s refers to the vapour pressure if the air was to be saturated with water vapour, and e_a to the actual vapour pressure in the air, each calculated as:

$$e_s = 0.6108 * e^{\frac{17.27 * Tair}{Tair+237.3}} \quad \text{Eq. 6}$$

$$e_a = 0.6108 * e^{\frac{17.27 * T_d}{T_d+237.3}} \quad \text{Eq. 7}$$

The gridded climate data was resampled to match the cell alignment of the S2 data using bilinear interpolation. A 50x50 pixel window of the modelled daytime values of VPD, PAR and T_{air} were extracted and compared to in situ measurements to assess the degree of uncertainty using ordinary least squares (OLS) regression, giving a R^2 of 0.80, 0.71, and 0.81, respectively (Figure 3). The modelled data was thereafter used as input to the GPP models (see section 4.2.4).

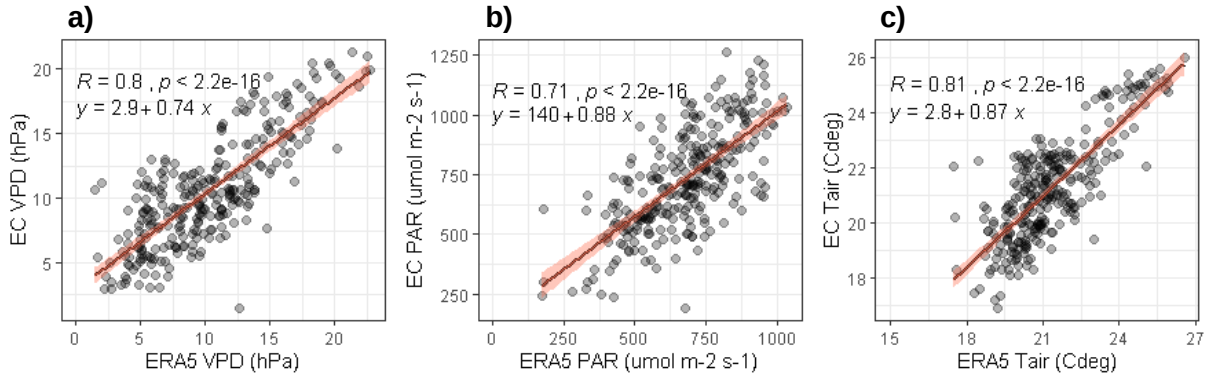

Figure 3: Regression statistics between modelled ERA5 data and in situ data for a) Vapour pressure deficit (VPD), b) Photosynthetically active radiation (PAR), and c) Air temperature (T_{air}). The points are made transparent to indicate the point density.

4.1.4 MODIS GPP

Being a commonly used GPP product and the only one available over Kenya, the MOD17A2H GPP v.6.0 product (MOD17) was compared to the developed models. The MOD17 GPP product is derived as:

$$GPP = 0.45\ x\ SW_{rad} x\ fAPAR\ x\ LUE_{max}\ x\ LUE_{VPD}\ x\ LUE_{Tmin} \quad \text{Eq. 8}$$

where, SW_{rad} is the incoming shortwave radiation, and LUE_{VPD} and LUE_{Tmin} are environmental stress scalars for VPD and temperature, respectively. The LUE terms are empirically determined and defined in a biome-specific look-up-table (BLUP), while climate observations stem from the National Center for Environmental Prediction Department of Energy (NCEP-DOE) reanalysis II data, represented in 1.9ºx1.9º grids. The fAPAR comes from MODIS fAPAR/LAI estimates, calculated in the MOD15 product. The resulting GPP is an 8-day cumulative composite expressed in kg C m^{-2} with a 500 m spatial resolution.

Data between January 1st and December 27th from the pixel centring the location of the flux tower was downloaded using the online tool AppEEARS from NASA's Earth Science Data and Information System (ESDIS) project.

4.1.5 Land cover map

The Copernicus Global Land Service (CGLS) dynamic LC map over Africa collection 2.02 was used for delineating LC types of cropland, grassland, open forest and shrubs, where the latter three classes were selected as approximately representing a savannah landscape mosaic. The dataset is derived for the epoch 2018 and has a 100 m spatial resolution. The main inputs are PROBA-V satellite observations. The classification follows the UN-FAO LC Classification System (LCCS) and the definitions for the used classes are seen in Table 2. Figure 4 shows the LCs within the study area. Although the dataset has not yet been formally validated, the global map from the same collection yielded an overall accuracy of 80% based on 21,600 validation points (Buchhorn et al. 2019).

Table 2: The definitions of the land cover classes used to separate GPP. The definitions follow UN-FAO LCCS and are taken from the user manual by Buchhorn et al. (2019).

Classification name	Definition
Cropland	*Lands covered with annual crops, assuming a bare soil period after harvest.*
Herbaceous vegetation (grassland)	*Plants without persistent stem/shoots above ground and lacking definite firm structure. Tree or shrub cover < 10 %.*
Open forest, unknown type	*Top layer consists of trees with 15-70 % coverage. Second layer consists of mixed shrubs and grassland.*
Shrubs	*Woody perennial plants with persistent and woody stems, without a defined main stem and <5 m tall. The shrub foliage can be either evergreen or deciduous.*

4.1.6 Agro-climatic zone map of Kenya

Since croplands are generally found in the sub-humid regions with higher potential yield, the agro-climatic zone (ACZ) map of Kenya was used to exclude sub-humid areas from the analysis and thus reduce the bias in the comparison between land cover classes. The map was created by the Kenya Soil Survey in 1981 and accessed as a shapefile layer from ICPAC geoportal (http://geoportal.icpac.net/). The data provides an overview of areas climatically suitable for different land use alternatives, with special focus on agriculture.

The ACZs are based on moisture availability and temperature. The moisture availability classes are defined based on the ratio of the measured MAP and the calculated mean annual evaporation. The temperature zones are based on the mean of the average maximum and minimum temperature. The combination of moisture availability and temperature make up an ACZ. However, only the former was used for the final analysis, as it was found that stratifying based on temperature added little additional information on the difference between the LC classes; the term ACZ will be used hereon, though actually referring to the moisture availability classification solemnly. The zones IV (semi-humid to semi-arid) and V (semi-arid) were selected for the analysis, while the arid zone (VI) was excluded since it only entailed 5 km^2 of cropland.

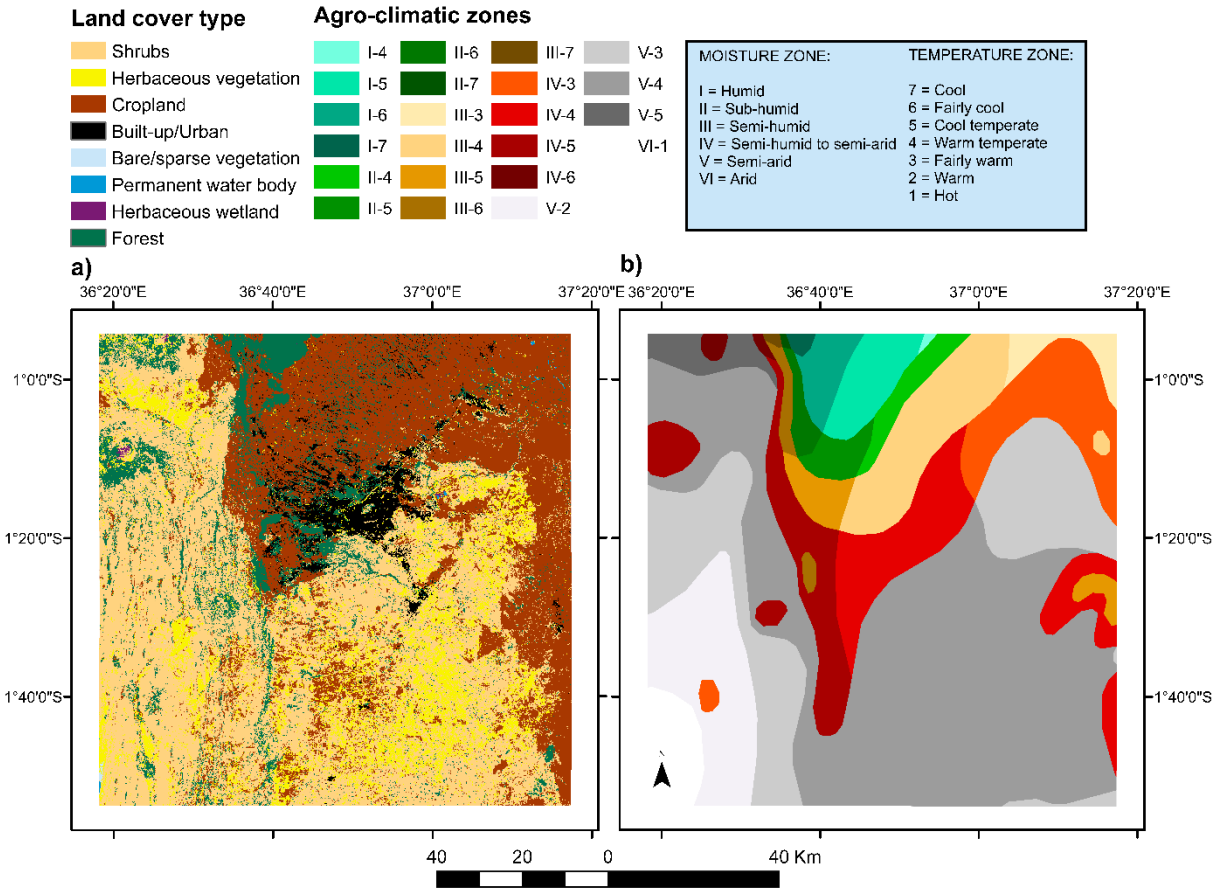

Figure 4: The data used for separating GPP into classes. a) The land cover type map, and b) the agro-climatic zone map, where the roman numbers stand for the moisture zone, and the Arabic number for the temperature zone. The explanation for the letters are found in the blue box in the figure.

4.2 Methods

Sections 4.2.1 and 4.2.2 detail the methods used to pre-process S2 and EC data, while the subsequent sections each will present the techniques used to answer the three aims of this thesis.

4.2.1 Sentinel-2 pre-processing

The temporal dimension of a vegetation index can be used as a descriptive of the phenological stages of the canopy layer. The level of greenness should conform to a somewhat smooth seasonal variation, however, noise from clouds, geometric distortion, anisotropic reflectance and other artefacts may lead to inaccurate, non-biophysical signals to be present in the data series. To resolve this, time series smoothing in TIMESAT v.3.3 (Eklundh and Jönsson 2015) was used. A Savitsky-Golay function was applied, using a data range from 0-1, window size=4, and envelope iterations and adaptation strength set to 3.

The processed output from TIMESAT was interpolated to derive daily values of EVI2 between the 1 January 2019 and 31 December 2019, using linear interpolation between the images. This step was performed after processing in TIMESAT to reduce the computational time. For calibrating and evaluating the GPP model, the median values of 50x50 pixels centred on the location of the EC tower were extracted. The extracted pixel window of the processed S2 data had a R^2=0.76 to spectral tower EVI2 and was disturbed by persistent cloud cover in the beginning and end of the growing season, displayed as dips in the S2 data in comparison to the spectral tower data (Figure 5). After the comparison, spectral tower EVI2 data was not included in further analysis.

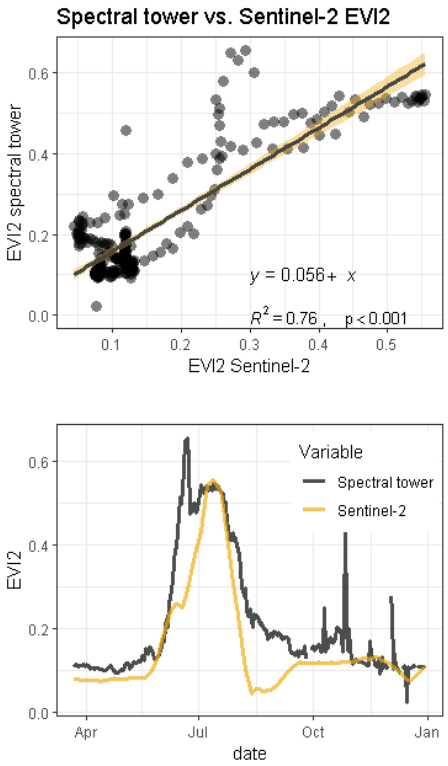

Figure 5: Comparison between ground-based spectral tower EVI2 and the processed Sentinel-2 (S2) EVI2 to assess the degree of accuracy of the S2 data. The dips in the S2 data denote cloud-cover disturbances, while peaks in the spectral tower data between October and December are faulty measurements.

4.2.2 Gap-filling and partitioning of flux data

Gap-filling and partitioning of NEE into GPP and R_{eco} were carried out using the online tool developed by Wutzler et al. (2018). For partitioning, the daytime method following Lasslop et al. (2010) was chosen, which uses daytime measurements (where global radiation > 4 W m^{-2}) and fits a hyperbolic light response curve to NEE. The curve is modified to account for sensitivity of R_{eco} to temperature (Lloyd and Taylor 1994), and the VPD limitation of photosynthesis. Prior to partitioning, data of insufficient turbulence (u*) was removed (in total 16% of the measurements) to avoid biased measurements (Papale et al. 2006).

Missing values were assigned the average value under similar environmental conditions within a 7-day time window. If no similar values were present (i.e. deviation by more than 50 W m^{-2} for global radiation, 2.5 °C for T_{air}, and 5.0 hPa for VPD), the window size was increased until the values could be filled.

4.2.3 Exploring the relationship of C to environmental variables

The first aim was achieved by using in situ micro-meteorological data and modelled SM from the ERA5 dataset in a regression tree analysis from where the variable importance (VI) score was derived. The influence of abiotic factors (in situ VPD, T_{air}, PAR and modelled SM) on in situ GPP was evaluated using a regression tree analysis based on the R package *caret*, which uses the CART (classification and regression tree) algorithm and calculates VI scores as part of the output. The method bagging (bootstrap aggregating) was used to grow 50 trees, each representing a possible scenario based on a new sub-sample of data. The method calculates the total amount that the total sum of squared estimate of errors (SSE) is decreased by splits over each variable, with the average taken over all 50 trees. Thus, the derived VI score is the relative average reduction in the SSE compared to the most important variable, and is scaled from 0 to 100, where 100 defines the most important variable. Along with VI, the RMSE and R^2 were calculated.

4.2.4 Model creation and statistical analysis

To answer the second aim, two non-linear asymptotic model and several multiple linear regression models were created and together with the MOD17 GPP product compared to in situ GPP. All models were created using S2 EVI2 data and gridded climate data. A non-linear modelling method was selected since the response of APAR ($EVI2 * PAR$) was found to be non-linearly related to GPP. To test a linear model including the importane of environmental constrains on GPP analysed in 4.2.1, a modelling method based on stepwise multiple linear regression analys was also applied. The two following sub-sections will descibe their derivation.

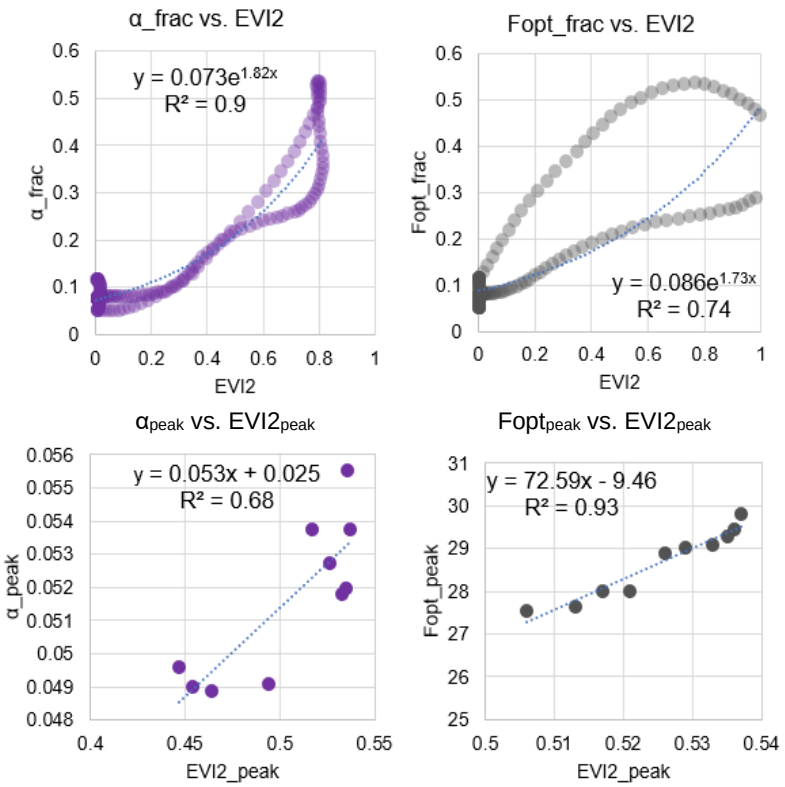

Figure 6: The regression functions used to fit α_{frac} and $Fopt_{frac}$ to S2 EVI2 (upper plots), and α_{peak} and $Fopt_{peak}$ to S2 $EVI2_{peak}$ (lower figures). The derived functions and their respective coefficients were thereafter used in the GPP model.

Fitting a non-linear light-response model

The light-response model used in this study follows the principals of Tagesson et al. (2017). The model uses the initial slope of the light-response curve (α (μmol CO2 m^{-2} s^{-1})) and the maximum photosynthetic rate at light saturation (Fopt (μmol CO_2 m^{-2})) to estimate GPP, based on a non-linear Mitscherlich asymptotic curve:

$$GPP = -Fopt * \left(1 - exp^{\frac{-\alpha - PAR}{Fopt}}\right) \qquad Eq.\ 9$$

where the Fopt and α are estimated from EVI2 using regression analysis to in situ values of Fopt and α. To derive Fopt and α from the EC data, eq. 9 was fitted to daytime in situ GPP using a 7-day running window and a 1-day time step. The peak values of S2 EVI2 ($EVI2_{peak}$), Fopt ($Fopt_{peak}$) and α (α_{peak}) were calculated as the peak value of a 14-day running mean. By dividing the values of EVI2, Fopt and α by their respective peak value, time series of fractions were derived ($EVI2_{frac}$, $Fopt_{frac}$ and α_{frac}).

To relate $Fopt_{frac}$ and α_{frac} to EVI2, exponential regression functions were used. To link α_{peak} and $Fopt_{peak}$ to EVI2, the ten highest values of α and Fopt were fitted against peak values of EVI2 using linear regression. The resulting regression functions can be seen in Figure 6. By connecting in situ Fopt and α to EVI2, the EVI2 was used to derive Fopt and α such as:

$$Fopt = Fopt_{peak} * Fopt_{frac} = \left(k_{Fopt} * EVI2_{peak} + m_{Fopt}\right) * \left(n_{Fopt} * e^{\left(l_{Fopt} * EVI2_{frac}\right)}\right) \qquad Eq.\ 10$$

$$\alpha = \alpha_{peak} * \alpha_{frac} = \left(k_{\alpha} * EVI2_{peak} + m_{\alpha}\right) * \left(n_{\alpha} * e^{\left(l_{\alpha} * EVI2_{frac}\right)}\right) \qquad Eq.\ 11$$

where *k* and *m* are slopes and intercepts for the linear equation, respectively, and *l* and *n* are slopes and intercepts for the exponential equation, respectively. The Eq. 10 and 11 were inserted in Eq. 12, and the coefficients derived from the linear and exponential regressions were finally used to calculate GPP as:

$$GPP = -\left(k_{Fopt} * EVI2_{peak} + m_{Fopt}\right) * \left(n_{Fopt} * e^{\left(l_{Fopt} * EVI2_{frac}\right)}\right) \qquad Eq.\ 12$$

$$* \left(1 - e^{\left(\frac{-\left(k_{a} * EVI2_{peak} + m_{\alpha}\right)\left(n_{\alpha} * e^{\left(l_{\alpha} * EVI2_{frac}\right)} * PAR\right)}{\left(k_{Fopt} * EVI2_{peak} + m_{Fopt}\right)\left(l_{Fopt} * EVI2 + n_{Fopt}\right)}\right)}\right)$$

To see wheter the parameterised coefficients from Tagesson et al. (2017) for the Sahel were applicable to this region, too, these were also tested in a separate model. The coefficients used in the final two models can be found in Table S1.

Multiple linear regression models

The data used in the statistical analysis (i.e. the calibration data) was selected between the 1st of May 2019 and the 30rd of August 2019, representing the period just before and after the growing season, where GPP is assumed to be related to environmental variables.

A stepwise multiple linear regression (MLR) was used to construct a GPP model using S2 EVI2 and modelled VPD, T_{air}, SM and PAR (Figure 7) as possible explanatory variables. To test for multicollinearity, the variance inflation factor (VIF) for all variables were calculated prior to the analysis. The results indicated that none of the factors were showing multicollinearity (VIF

< 10). Moreover, the VPD was found to be leading GPP, thus creating a positive relationship between VPD and GPP. To resolve this, cross correlation was used to obtain an approximate lag time of VPD, by choosing the lag with the highest correlation and significant p-value. The lagged VPD was thereafter used as input in the MLR model. The stepwise MLR method randomly adds and discards variables until attaining a model with the highest fit while penalising adding more independent var-iables. The Bayesian Infor-mation Criterion (BIC) (Eq. 13) and R^2 were used to select a suitable MLR model as following:

$$BIC = -2\,ln(L) + q\,ln(n) \qquad \text{Eq. 13}$$

where L is the maximized value of the likelihood function of the model, and n is the number of observations.

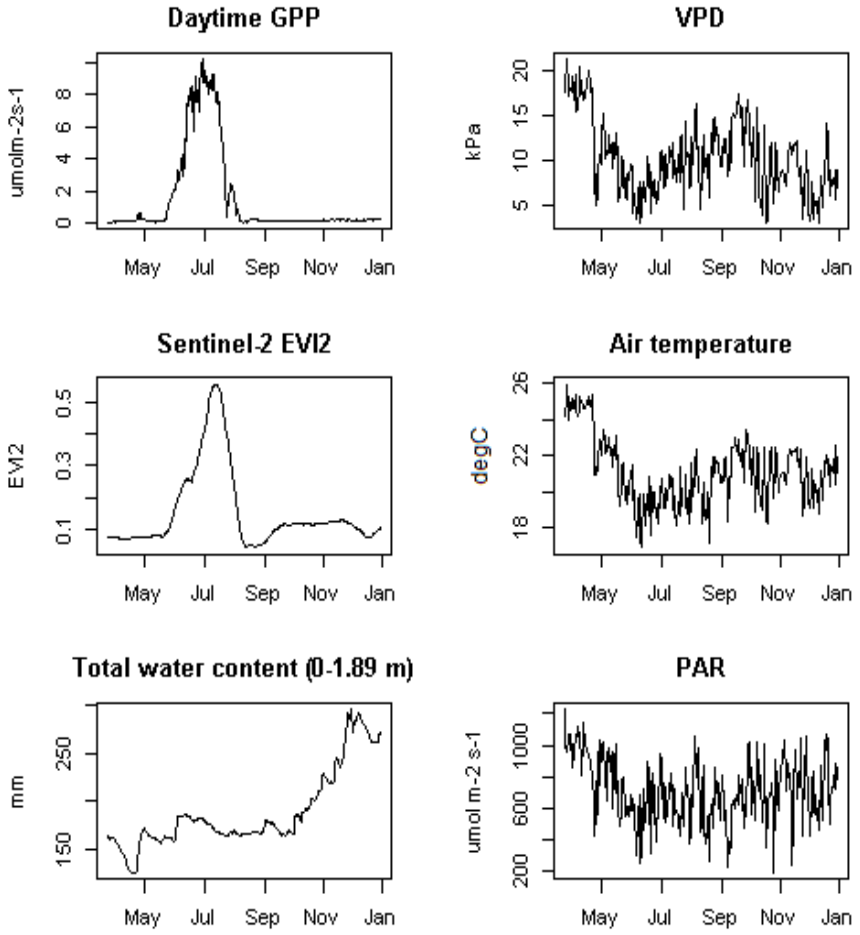

Figure 7: The data used as possible explanatory variables in the GPP model, and in situ GPP (upper left). The climatic variables are derived from the ERA5 dataset. VPD = Vapour pressure deficit. PAR = Photosynthetically active radiation.

Deciding on a model for upscaling

As a measure of performance, OLS regression between each model, based on a 50x50 window size of gridded climate data and S2 data, to in situ GPP was used. Based on these results, the model with the highest R^2 and lowest RMSE over the full measurement period was used as criteria for selecting a model for upscaling over the full S2 tile.

4.2.5 Upscaling and stratification

The upscaling was performed using TerrSet v. 18.3 (Clark Labs, Clark University, USA), by using the attained relationship for the selected model (see results 5.3) found in section 4.2.4. The cumulative annual modelled GPP was calculated for each pixel and compared for each of the four selected LC classes. To reduce the bias of croplands being situated in the more humid, productive zones, only the values falling within zone IV and V were used in the analysis. Thus, the GPP was stratified based on both LC class and ACZ, creating in total 8 different classes.

To see whether the difference in GPP is related to the length of the growing seasons, all cells with GPP > 2 g C m^{-2} d^{-1} were summed over the year as to approximate the start and end of the growing season. The threshold of 2 g C m^{-2} d^{-1} was selected by visually comparing time series of modelled daily GPP of random pixels from croplands and natural drylands. Figure 8 shows an example of ten time series. The Pearson's correlation coefficient was calculated between the length of the growing season and the annual GPP using a 9x9 pixels moving window. The resulting maps of the length of the season and spatial correlation were stratified the same way as the annual GPP.

The comparison between the classes was drawn using the zonal statistics tool in ArcGIS 10.5 (ESRI Inc., California, USA) and visually using notched boxplots in R. Notches are useful as they can give a rough estimate whether there is a significant difference between the medians of groups; if the notch-intervals do not overlap, the two groups being compared are approximately different at the 5% level (Chambers et al. 1983). To each boxplot, a sample size of 100,000 random pixels was used.

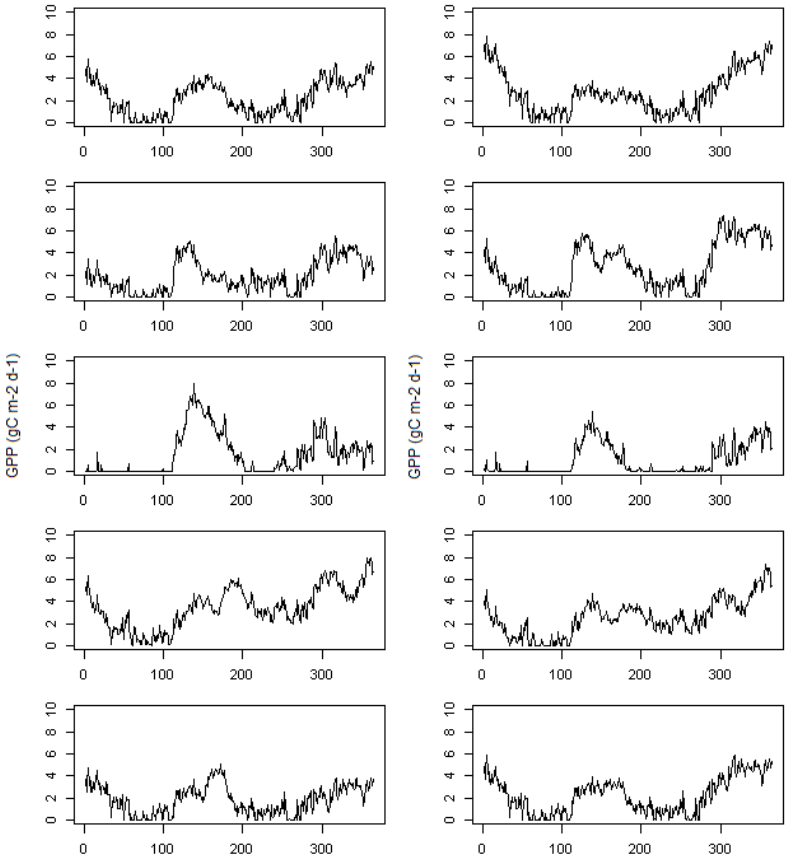

Figure 8: Example of 10 random pixels of savannah land cover character or cropland, used for defining a threshold for separating the growing/non-growing season based on modelled GPP.

5. Results

This section is divided according to the aims. It will start by presenting results from in situ measurements of C fluxes and the influence of climatic variables on GPP. The second section shows the performance of the different models and the MOD17 GPP product to measured GPP. The third section summarises the upscaling results of the chosen model by disintegrating the spatial patterns seen in GPP by LC class and ACZ.

5.1 Cropland C variability and controlling factors

The growing season at Ausquest farm started mid-May, with a peak in GPP of 10.3 g C m^{-2} d^{-1} on the 29th of June and ended in the beginning of August (Figure 9a). The R_{eco} had the largest cumulative flux for the measurement period (2019-03-23 to 2019-12-30) of 660 g C m^{-2}, while cumulative GPP was 438 g C m^{-2}. The field had a net loss of 170 g C m^{-2} during the year 2019 (denoted as a positive NEE) (Figure 9b).

The VPD has the largest influence on GPP during the growing season, followed by SM. The T_{air} has little influence, while PAR has no profound impact (Table 3). Values of GPP above 1 g C m^{-2} d^{-1} occurs when SM is above 160 mm and when VPD is under 15 hPa.

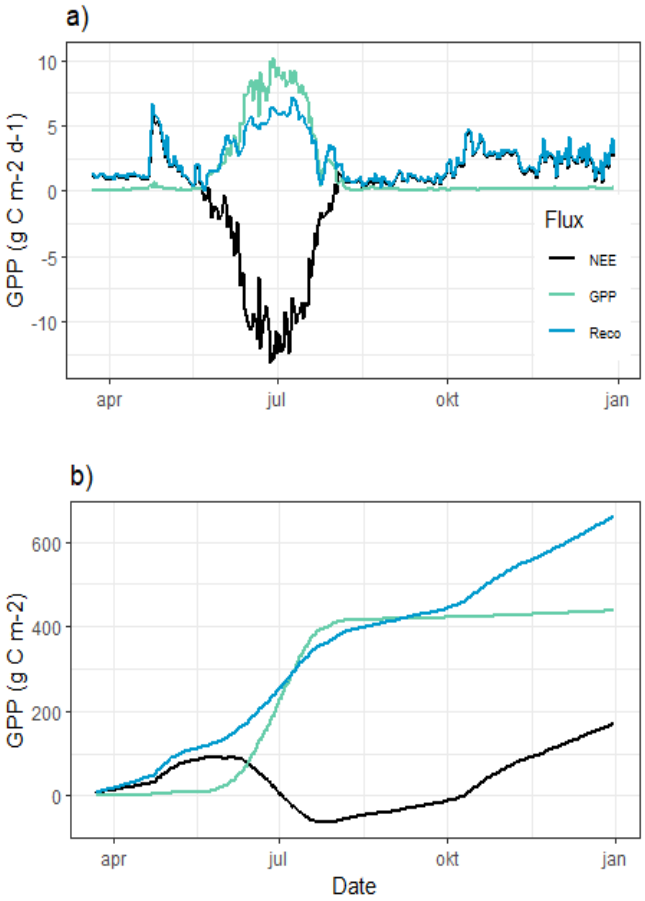

Figure 9: Fluxes of NEE (black), GPP (light green) and Reco (blue) plotted as a) time series of daily values, and b) Cumulative values for the measurement period (2019-03-23 to 2019-12-30).

Table 3: The results from the regression tree. VI = Variable importance score.

Variable	**VI**
VPD	100
SM	87
T_{air}	11
PAR	0
RMSE	**2.1 g C m^{-2} d^{-1}**
R^2	**0.65**

Response variable = GPP

5.2 Modelling GPP

Various RS driven models were tested against EC GPP prior to upscaling in order to find the most accurate way to model GPP. These included MLR models, two non-linear (hereby referred to as NL) models based on the relationship between APAR and GPP, and the MOD17 GPP product. The model with the highest R^2 and lowest RMSE was thereafter used for upscaling in section 5.3.

5.2.1 The NL models

Two NL models were tested; one calibrated to the flux measurements (model A), and one based on the derived values by Tagesson et al. (2017) (model B). Model A underestimates growing season GPP while overestimating the non-growing season, resulting in a low fit (R^2=0.42 and RMSE= 2.36 g C m^{-2} d^{-1}) (Figure 10a). As seen in Figure 10b, the Fopt and α of model A are low compared to EC Fopt and α. Model B captures the end of the season and the non-growing season well, but misses the first half of the growing season, which lowers the overall fit (R^2=0.49, RMSE=2.09 g C m^{-2} d^{-1}) (Figure 10a). Peak values of Fopt and α of model B were slightly higher compared to in situ values (Figure 10b), with prevailing underestimation in the first half of the season. Since the modelled Fopt and α are derived from the EVI2, modelled Fopt and α are closely following the pattern of the S2 data. As seen in Figure 10c, the EVI2 is lower compared to GPP for the start of the season, when cloud cover was persistent.

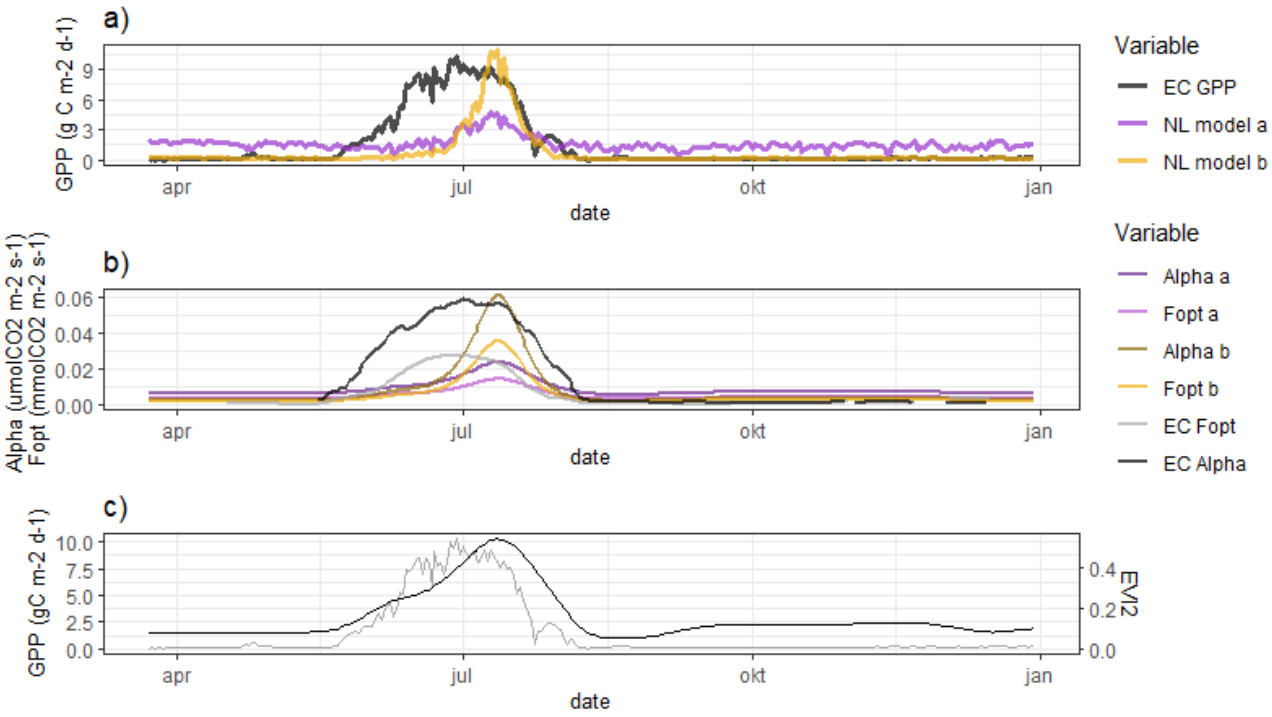

Figure 10: Time series of in situ measurements (EC), the NL model A (purple), and the NL model B (yellow), where plot a) shows daily GPP, b) shows the 14-day running mean Fopt and α (alpha), and c) shows EC GPP (grey) and the S2 EVI2 (black).

5.2.2 The MLR models

Table 4 gives a summary of the MLR model performances for the growing season, which was the period used for calibrating the models. The MLR model with the highest fit and lower BIC for the growing season included EVI2, SM, VPD and Tair (R^2=0.86, BIC = 436), the same variables indiated as important in the regression tree analysis (Table 3). However, this model resulted in a poor fit for the full validation period (R^2 = 0.1) (Figure 11), caused by the amplified influence of SM on GPP during the second rain period in October-December, when SM was replenished while management practises controlled growth. Therefore, a model without the inclusion of SM as a explanatory variable was neccessary. This lowered the fit over the growing season, but prevented some of the overestimation during the non-growing season.

Table 4: Results for the multiple linear regression models with the best fit (R^2) and lowest BIC for just before and after the growing season (n = 104). SM = soil moisture 0-189 cm. lagVPD = lagged VPD with 30 days. All parameters have a p-value < 0.001 if not denoted otherwise.

Independent variables	R^2	RMSE	BIC
EVI2 + lagVPD + SM + Tair	0.87	1.25	436
EVI2 + lagVPD	0.74	1.81	502
EVI2 + lagVPD + Tair*	0.74	1.80	507

Response variable = GPP, * = p-value > 0.1

After excluding SM from the regression analysis, the MLR model using EVI2, VPD, and T_{air} as explanatory variables (referred to as $MLR_{VPD,Tair}$) yielded the best fit for the full measurement period (R^2=0.65, RMSE = 2.06 g C m^{-2} d^{-1}) (Table 5).

The $MLR_{VPD,Tair}$ model generally follows the seasonality of the EC GPP data (Figure 11). The largest discrepancies between modelled GPP and EC GPP are seen during the two rain periods (March-May and October-December), when the model overestimates GPP with about 3 g C m^{-2} d^{-1}, and in the beginning of the growing season, where the model underestimates GPP.

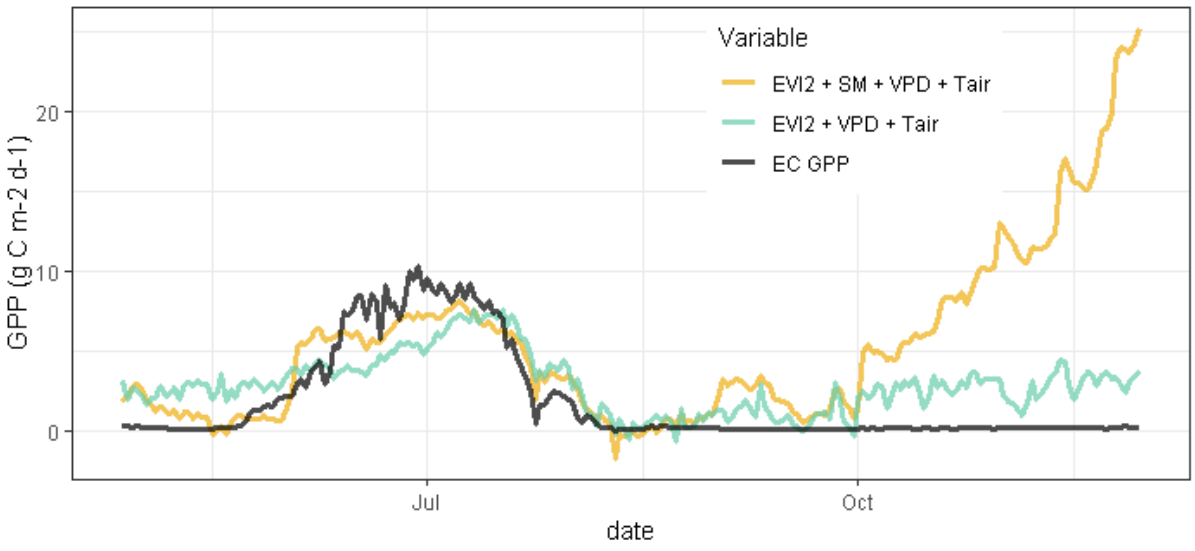

Figure 11: The results of including vs. excluding soil moisture (SM) in the multiple regression model. The yellow line depicts the model where SM is included, whereas the green line is the $MLR_{VPD,Tair}$ model.

5.2.3 Comparison of models

Table 5 gives a summary on the performance of all models and the MOD17 product. The $MLR_{VPD,Tair}$ model depicted in Figure 11 explains most of the variance compared to the derived models. In comparison to the MOD17 product, the $MLR_{VPD,Tair}$ model explains 67% (8-day cumulative GPP) and has a slope and intercept closer to 1 and 0, respectively (Figure 12a-b). Both the MOD17 product and $MLR_{VPD,Tair}$ model overestimate GPP during the non-growing period and underestimate GPP during the growing period, but the $MLR_{VPD,Tair}$ model captures the intermidate drier months in April and September slightly better than MOD17 (Figure 12c). Overall, the $MLR_{VPD,Tair}$ model had the best fit to EC GPP compared to both the derived models and the MOD17 product.

Table 5: The performance of all tested models and the MOD17 product against EC GPP over the full measurement period. n = 283, unit g C m^{-2} d^{-1}.

Model	R^2	RMSE
NL: Model A	0.42	2.36
NL: Model B	0.49	2.09
MLR: EVI2 + VPD + T_{air} + SM	0.1	6.72
MLR: EVI2 + VPD + T_{air}	0.65	2.03
MLR: EVI2 + VPD	0.64	2.04
MOD17 GPP	0.55*	17.1*

*8-day cumulative value, n=45

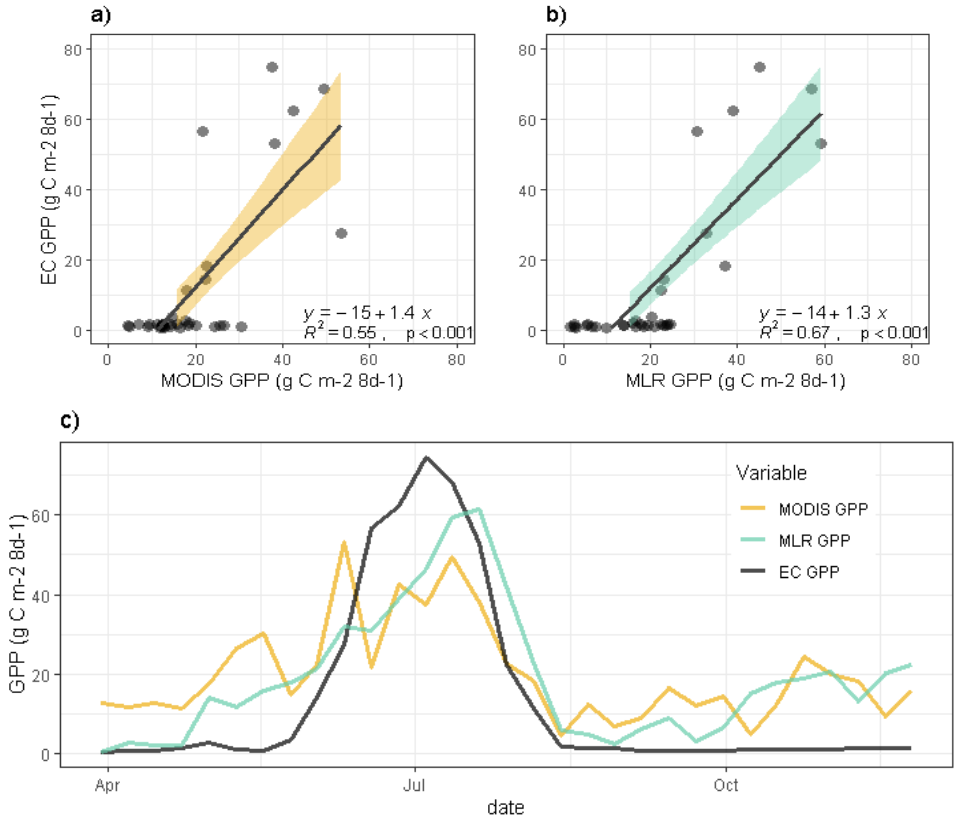

Figure 12: A comparison between 8-daily cumulative in situ GPP and a) the best-performing derived model, the $MLR_{VPD,Tair}$ model (MLR GPP), and b) MOD17 GPP (MODIS GPP), and c) Time series between in situ GPP and the two models, respectively. n = 45.

5.3 Upscaling GPP

Based on the results in 4.2, the $MLR_{VPD,Tair}$ model was used for upscaling GPP over the whole study area:

$$GPP = 4.7 + (12.7 * EVI2) + (-0.26 * VPD) + (-0.07 * T_{air}) \quad \text{Eq. 14}$$

Figure 13 shows the cumulative annual GPP for year 2019. The spatial pattern is closely following the ACZ, with higher annual GPP in the northern Kenyan highlands and eastern parts of the study area, and lower values in the semi-arid to arid southern and western parts. The western side of the cropland at Ausquest farm (location of flux tower) has a lower GPP

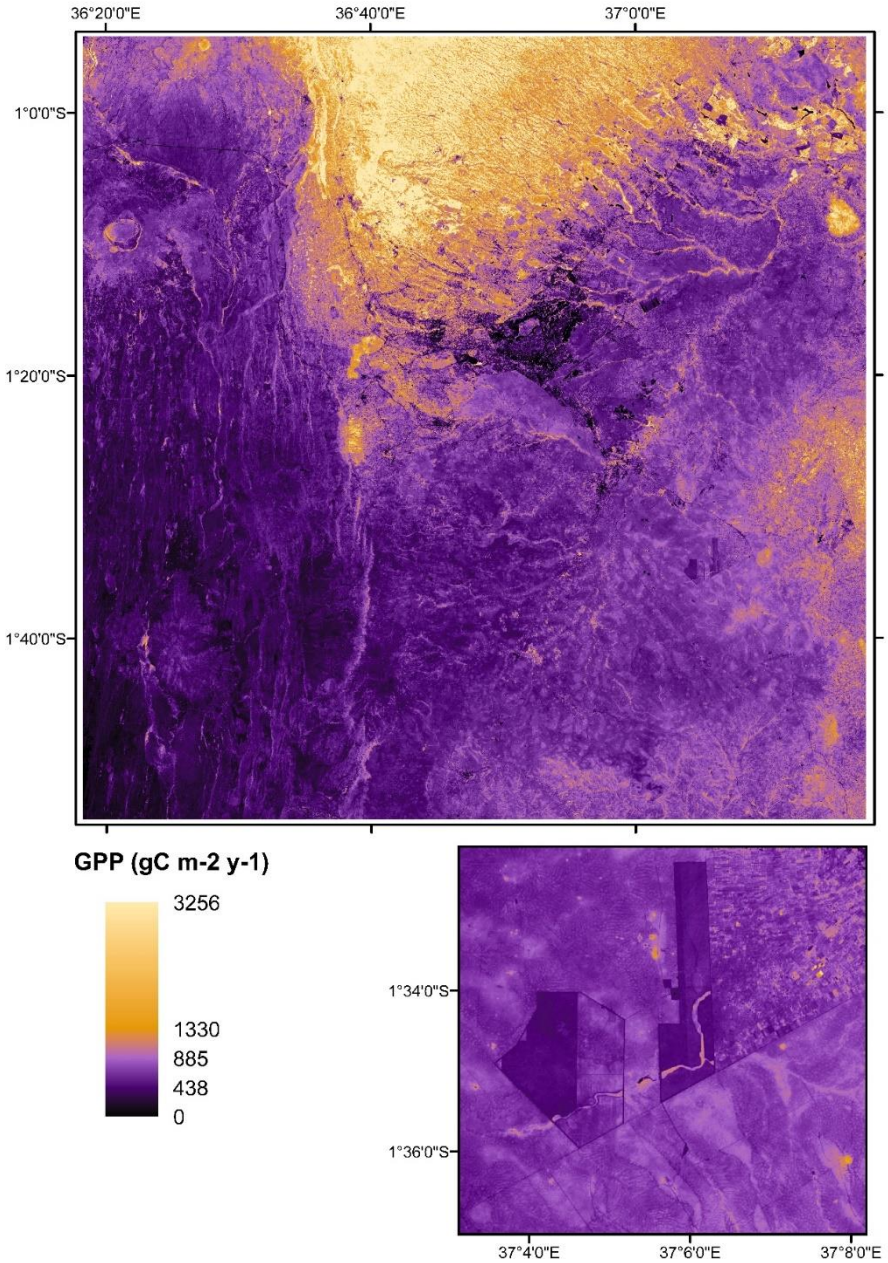

Figure 13: The spatial distribution of cumulative annual GPP (g C m^{-2} y^{-1}) for year 2019 within the study area. The lower, smaller image depicts the Ausquest farm and its surrounding savannah landscape south and north of the field, and small-scale croplands in the eastern corner.

compared to its surrounding, while the cropland's eastern side has the same GPP as the adjacent savannah landscape (Figure 13).

Table 6: A summary of the mean and standard deviation (sd) of the annual GPP (g C m^{-2} y^{-1}), the number of growing days (GD), defined as number of days where GPP > 2 gC m^{-2} d^{-1}, and areal extent for each LC class and zone. The largest values in each comparison is marked with bold text.

	Both zones			**Semi-humid to semi-arid (zone IV)**			**Semi-arid (zone V)**			**Mean diff. between zones**	
	GPP, mean (sd)	**GD**	**Area (km^2)**	**GPP, mean (sd)**	**GD**	**Area (km^2)**	**GPP, mean (sd)**	**GD**	**Area (km^2)**	**GPP (%)**	**GD (%)**
Cropland	**959 (301)**	**209 (60)**	2317	**1085 (340)**	**237 (62)**	**1079**	834 (263)	115 (52)	1238	**-23%**	**-24%**
Grassland	751 (269)	165 (60)	1222	792 (325)	174 (72)	101	711 (212)	94 (42)	1121	-10%	-10%
Shrubland	735 (248)	164 (57)	**5110**	782 (246)	176 (58)	613	687 (250)	93 (47)	**4497**	-12%	-13%
Open Forest	918 (298)	206 (63)	669	999 (307)	224 (61)	243	**838 (289)**	**187 (65)**	426	-16%	-16%

Croplands have a higher mean annual GPP compared to the other classes, except for open forest in zone V (Table 6), and a consistently higher median, though this difference is less evident in comparison to open forest in zone V (Figure 14a). Aggregating the classes grassland, shrubland and open forest to resemble a savannah mosaic landscape of increasing woodiness, thus accounting for the areal coverage of each class, yields a lower annual GPP compared to croplands in both zones (Figure 14c).

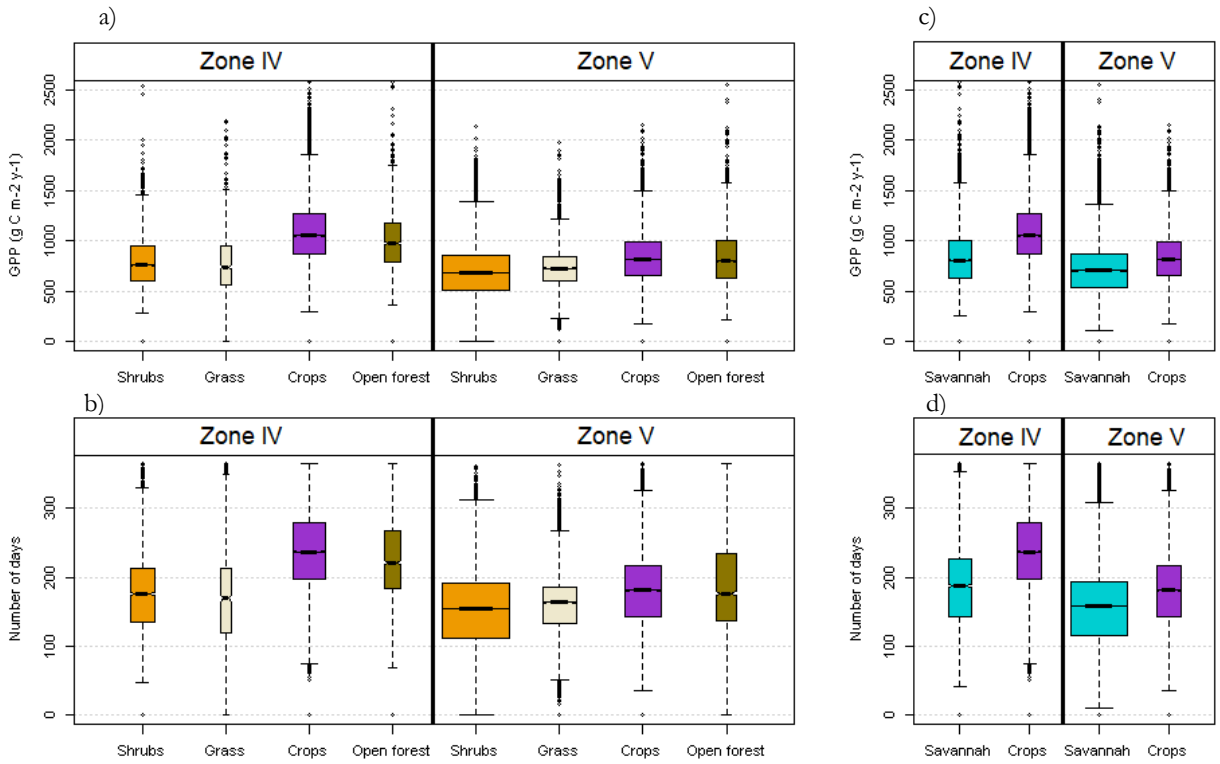

Figure 14: For each of the four land cover classes shrubland (orange), grassland (beige), cropland (purple) and open forest (green), notched boxplots of a) annual GPP for 2019 per unit area, and b) number of days where GPP > 2 g C m^{-2} d^{-1}, are shown. Plot c) and d) shows the values if merging shrubland-grassland-open forest classes to represent a savannah landscape mosaic (blue) vs. cropland. The width of the plots indicates the sample size. The limit of samples to use for plotting is set to 100,000 pixels.

Table 7: The average and standard deviation (within brackets), of Pearson's correlation coefficient between annual GPP and growing season for each land cover class within each of the zones semi-arid to sub-humid (IV) and semi-aid (V).

	R	
	Zone IV	**Zone V**
Cropland	0.86 (0.13)	0.84 (0.13)
Grassland	0.87 (0.12)	0.82 (0.15)
Shrubland	0.87 (0.12)	**0.88 (0.12)**
Open Forest	**0.88 (0.12)**	**0.90 (0.09)**

To determine the influence of the length of the seasons on annual GPP, the number of growing days (defined as GPP > 2 g C m^{-2} d^{-1}) was calculated (Figure 15). The resulting difference between the groups was closely following the pattern of annual GPP (Figure 14b). Croplands on average had the greatest number of growing days in zone IV, while open forests have more days in zone V (Table 6). The correlation between annual GPP and growing days were highest for open forest (R=0.90) and shrubland (R=0.88) in zone V, and lowest for cropland (R=0.84) and grassland (R=0.82) in zone V (Table 7).

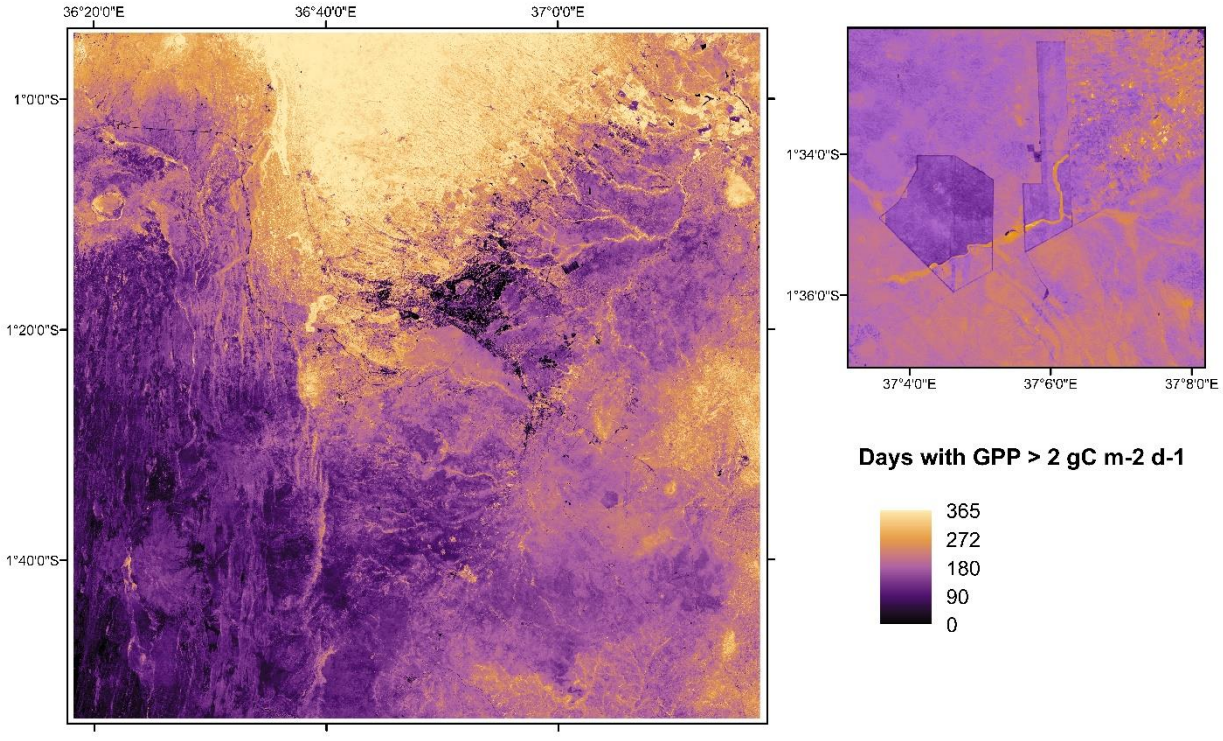

Figure 15: The number of days where GPP > 2 g C m^{-2} y^{-1} (left). The square depicts the results for the Ausquest farm and its surroundings (right).

6. Discussion

Based on a unique dataset of continuous flux measurements in Kenya, the aims of this study were to determine the net C flux and climatic factors governing GPP at the Ausquest farm, to use this as a first opportunity to upscale GPP to the surrounding region, and to assess the regional difference in GPP between cropland and natural semi-arid land.

This chapter is divided into four parts. The initial three sections are devoted to each of the aims of the thesis, finalised with a section discussing limitations and future studies.

6.1 Cropland C balance and environmental constrains

The studied cropland was a source of C between mid-March 2019 until the end of December 2019, indicating the importance of including respiration in LC change impacts on the C cycle. Previous studies using EC data have indicated that croplands in Europe and USA on average act as larger sinks of C compared to savannahs (Gilmanov et al. 2010), while Tagesson et al. (2016) in a study over the Sahel found human land management to rather reduce the C sink strength, attributed to the reduction in GPP. However, abiotic circumstances, crop type and management strategy impact the net flux of Kenyan croplands (Rosenstock et al. 2016; K'Otuto et al. 2013), consequently making it difficult to extrapolate these results to other agricultural fields. To better understand impacts of LCC on the C cycle and to adapt sustainable agricultural intensification practises in the Sub-Saharan region, which currently is very low (Kurgat et al. 2018), more studies on the C cycling of croplands on a country to continental scale is needed.

The results indicate that water availability is governing GPP at the site, and that both VPD and SM were important descriptors, in line with previous studies over the Sahel (e.g. Tagesson et al. 2017; Abdi et al. 2017). However, EC data over more years and locations is likely needed to fully capture the dependency of GPP to environmental factors, especially considering the large influence of land management on GPP at the Ausquest farm.

6.2 Model performances

6.2.1 Evaluation of the MLR model

Using an MLR yielded the highest fit to in situ GPP and captured the growing season better than the NL models and the MOD17 GPP product. In accordance to Abdi et al. (2017), it is a simple yet effective method for upscaling GPP over a semi-arid region, while taking advantage of higher-resolution RS data and climate data, more adapted to the study area. The MLR model performed better compared to the MOD17 GPP product in terms of fit and magnitude. The underestimation of growing season GPP in MOD17 was expected, and in agreement with observations form semi-arid Sahel (Tagesson et al. 2017; Abdi et al. 2019), and over croplands

in the USA (Xin et al. 2015; Wang et al. 2017). However, comparing the performance of the two models to more ecosystems would have added further value to the analysis.

The MLR model overestimated GPP over the rain periods, when VPD decreased, but land management suppressed GPP. This was exaggerated even more when adding SM from down to 189 cm in the model. Fitting the MLR model to more EC data, preferably also with two growing seasons, is likely to make the relationship between environmental variables, EVI2 and GPP more robust. However, as with the MOD17 GPP product, the underlying assumption of the method is that there is a biophysical explanation for variations in GPP. This makes both methods naturally flawed when applied on managed land. To incorporate a Boolean variable describing the reliance of GPP on land management rather than water availability when EVI2 is sufficiently low, could help improving the modelling performance of an MLR model on croplands, especially those having only one growing season in a region with two rain periods.

6.2.2 Evaluation of the NL model

Since the NL method does not incorporate water availability as a factor describing GPP, the NL models did not have the same issue of overestimating GPP in the non-growing season. However, the modelled growing season did not start until EVI2 was around 0.3 (Figure 10c), consequently making the NL models miss the cloud-covered beginning of the growing season. The MLR model did not show the same sensitivity to cloud disturbances, possibly because the MLR model includes variables not derived from EO data that can enhance GPP even though EVI2 is low, while cloud cover in the NL model rather reduces both PAR and EVI2. Though including VPD was proven to be a shortcoming of the MLR model in the non-growing season, it could have been of benefit for modelling GPP when the EVI2 was of low quality.

An explanation for the underperformance of model A is the lack of further parameterisation of the regression functions between EVI2, Fopt and α; in Tagesson et al. (2017), a bootstrap method was used to optimise the model fit to GPP. This step was not performed in this thesis, resulting in too low Fopt and α. Performing this step could have improved the magnitude of the modelled GPP, but considering the influence of EVI2 on the model results, further parameterisation is not likely to have improved the fit over the beginning of the season.

In Tagesson et al. (2017), it was noted that the NL model overestimated GPP over croplands (Millet farm). Though the results indicate that the modelled GPP follows in situ GPP quite well for the latter part of the growing season, there was a small overestimation present. Thus, it might be necessary to adjust the model parameters to better fit the GPP of croplands. However, overall, if cloud cover disturbances are reduced in the S2 data, the model parameters derived by Tagesson et al. (2017) seem to be applicable to the Kenyan region too.

6.3 Difference in carbon uptake

The third aim of this thesis was to use the derived model to determine whether annual GPP differ between croplands and natural semi-arid LC within the study area. The results show that

the annual GPP of croplands is larger than vegetation types typical for the savannah landscape in the semi-arid parts of the region, albeit the open forest class had a similar or slightly higher mean. The notched boxplot indicates that the classes shrubland and grassland have a significantly lower GPP compared to croplands, while the difference between cropland and open forest is harder to note (Figure 14a). If assuming that the open forest class is mostly resembling a savannah in accordance to Buchhorn et al. (2019), the results are generally in line with previous findings on a continental scale (Ardö 2015).

The studied field at the Ausquest farm had a noticeably lower GPP compared to its surrounding savannah landscape and small-scale farmlands, indicating that land management is important for understanding spatiotemporal changes in GPP and can impact the difference between LC classes. Furthermore, cropland annual GPP decreased the most between zone IV and V, implying climate is another factor determining the magnitude of the difference between croplands and savannah landscape mosaic classes. Consequently, the difference between LC classes is likely to depend on local-to-regional land management practices and climate, which has implications for both land management and C cycle studies.

Lastly, the results imply that the annual GPP of shrubland and open forests is more connected to the length of the seasons, while annual GPP of croplands and grasslands are more connected to the magnitude. These two groups of LC classes have a majority of evergreen vs. annual vegetation, thus being a possible explanation for the said difference. However, a larger analysis is needed to draw any further conclusions on this.

6.1 Limitations and improvements

Although expected to vary, the methods used in this thesis assume that the same relationships between GPP and environmental variables or EVI2 are valid for a plethora of ecosystems. Though inextricable with the lack of flux stations in Kenya and the surrounding countries, this introduces a high uncertainty in the modelled results. A clear improvement for both model calibration and validation would be to have access to more EC flux data.

Despite TIMESAT being a robust software to process RS time series, there is a clear limitation of using optical RS in cloud-prone areas for accurate GPP modelling. Incorporating EO data such as synthetic aperture radar (SAR) could help overcome these shortcomings.

The estimated difference in GPP is largely affected by not only the GPP model, but the assumption that there is a discrete *natural land* and *anthropogenic land* separation between them, and that the LCs within each ACZ have the same potential yield. In this thesis, it was implicitly assumed that the GPP of grasslands, shrubland and open forest in the study area were not heavily impacted by human management. However, pastoralism, shifting cultivation and wood harvest for domestic usage are commonly practised in the region (Kiruki et al. 2017). As stated by Williams et al. (2008), these factors are seldom considered, though likely to have an impact on the C balance. Furthermore, croplands are not random in space; contrarily, they are logically placed in locations of higher productivity potential, introducing inevitable bias when comparing GPP. As such, a time-series study of changes in GPP over areas where LC changes

have occurred might be necessary to fully be able to compare the classes and the adjacent change in GPP, as well as to see trends in GPP within the classes.

7. Conclusion

Overall, this study shows that it is possible to upscale GPP using a combination of EC GPP data, EVI2 from S2, and modelled climate data with adequate results. The VPD and SM were both important for understanding dynamics in GPP. The MLR model had the best performance (R^2=0.65, RMSE = 2.03 g C m^{-2} d^{-1}), albeit the quality of the S2 data, land management, and lack of EC data for calibration and validation reduced the possibility to derive a model of higher fit. The NL models were largely affected by cloud cover, and for future studies the S2 data should be further processed or accompanied by additive EO data.

This thesis has shown that the high-intensively cropped dryland farm at the Ausquest farm had a net release of 170 g C m^{-2} during the 9 months of measurement and a considerably lower GPP compared to its surroundings. However, over the full study area, croplands on average had a higher annual GPP and a longer growing season compared to grasslands and shrublands, and an approximately equal annual GPP and growing season to open forests. This difference is likely affected by land management and climate and can thus vary depending on the comparison. Based on these results, it is probable that a cropland expansion over this region would not considerably change annual GPP, although it is important to consider the different time period that C is stored in natural vegetation versus crops, and possible increased respiration rates resulting in a net release of C.

A reoccurring conclusion is that the paucity of EC stations and general data on environmental variables over Kenya, and (Eastern) Africa overall, is largely hampering the possibility for a robust analysis. More ground-truth data is crucial for future studies.

References

Abdi, A., N. Boke-Olén, D. Tenenbaum, T. Tagesson, B. Cappelaere, and J. Ardö. 2017. Evaluating Water Controls on Vegetation Growth in the Semi-Arid Sahel Using Field and Earth Observation Data. *Remote Sensing,* 9**:** 294. DOI: 10.3390/rs9030294

Abdi, A. M., N. Boke-Olén, H. Jin, L. Eklundh, T. Tagesson, V. Lehsten, and J. Ardö. 2019. First assessment of the plant phenology index (PPI) for estimating gross primary productivity in African semi-arid ecosystems. *International Journal of Applied Earth Observation and Geoinformation,* 78**:** 249-260. DOI: https://doi.org/10.1016/j.jag.2019.01.018

Ahlström, A., M. R. Raupach, G. Schurgers, B. Smith, A. Arneth, M. Jung, M. Reichstein, J. G. Canadell, et al. 2015. The dominant role of semi-arid ecosystems in the trend and variability of the land CO2 sink. *Science,* 348**:** 895. DOI: 10.1126/science.aaa1668

Allaby, M. 2006. *Grasslands*. Facts On File Inc.

Allen, R. G., L. S. Pereira, D. Raes, and M. Smith. 1998. Meteorological data. In *Crop evapotranspiration - Guidelines for computing crop water requirements*. Rome, Italy: United Nations, FAO.

Ardö, J. 2015. Comparison between remote sensing and a dynamic vegetation model for estimating terrestrial primary production of Africa. *Carbon balance and management,* 10**:** 8. DOI: 10.1186/s13021-015-0018-5

Ardö, J., M. Mölder, B. A. El-Tahir, and H. A. M. Elkhidir. 2008. Seasonal variation of carbon fluxes in a sparse savanna in semi arid Sudan. *Carbon balance and management,* 3**:** 7.

Beer, C., M. Reichstein, E. Tomelleri, P. Ciais, M. Jung, N. Carvalhais, C. Rödenbeck, M. A. Arain, et al. 2010. Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate. *Science,* 329**:** 834-838. DOI: 10.1126/science.1184984

Boke-Olén, N., V. Lehsten, J. Ardö, J. Beringer, L. Eklundh, T. Holst, E. Veenendaal, and T. Tagesson. 2016. Estimating and Analyzing Savannah Phenology with a Lagged Time Series Model. *PLOS ONE,* 11**:** e0154615. DOI: 10.1371/journal.pone.0154615

Bond, W. J., F. I. Woodward, and G. F. Midgley. 2005. The global distribution of ecosystems in a world without fire. *New phytologist,* 165**:** 525-538.

Brink, A. B., C. Bodart, L. Brodsky, P. Defourney, C. Ernst, F. Donney, A. Lupi, and K. Tuckova. 2014. Anthropogenic pressure in East Africa—Monitoring 20 years of land cover changes by means of medium resolution satellite data. *International Journal of Applied Earth Observation and Geoinformation,* 28**:** 60-69. DOI: https://doi.org/10.1016/j.jag.2013.11.006

Bronswijk, J. J. B. 1991. relations between vertical soil movements and water-content changes in cracking clays. *Soil Sci. Soc. Am.* , 55**:** 1220-12226.

Buchhorn, M., B. Smets, L. Bertels, M. Lesiv, N.-E. Tsendbazar, and L. Li. 2019. *PRODUCT USER MANUAL: Moderate dynamic land cover collection 100 m version 2*. Copernicus Global Land Operations. https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS1_PUM_LC100m-V2.0_I2.20.pdf (last accessed.

Bucini, G., and N. P. Hanan. 2007. A continental-scale analysis of tree cover in African savannas. *Global Ecology and Biogeography,* 16**:** 593-605.

Chambers, J. M., W. S. Cleveland, B. Kleiner, and P. A. Tukey. 1983. *Graphical methods for data analysis*. Wadsworth & Brooks/Cole.

Ciais, P., A. Bombelli, M. Williams, S. L. Piao, J. Chave, C. M. Ryan, M. Henry, P. Brender, et al. 2011. The carbon balance of Africa: synthesis of recent research studies. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* 369**:** 2038-2057. DOI: doi:10.1098/rsta.2010.0328

Eklundh, L., and P. Jönsson. 2015. TIMESAT: A software package for time-series processing and assessment of vegetation dynamics. In *Remote sensing time series*, 141-158 pp.: Springer.

Fensholt, R., I. Sandholt, M. S. Rasmussen, S. Stisen, and A. Diouf. 2006. Evaluation of satellite based primary production modelling in the semi-arid Sahel. *Remote Sensing of Environment,* 105**:** 173-188. DOI: 10.1016/j.rse.2006.06.011

Giannini, A., M. Biasutti, I. M. Held, and A. H. Sobel. 2008. A global perspective on African climate. *Climatic Change,* 90**:** 359-383.

Gilmanov, T. G., L. Aires, Z. Barcza, V. S. Baron, L. Belelli, J. Beringer, D. Billesbach, D. Bonal, et al. 2010. Productivity, Respiration, and Light-Response Parameters of World Grassland and Agroecosystems Derived From Flux-Tower Measurements. *Rangeland Ecology & Management,* 63**:** 16-39. DOI: https://doi.org/10.2111/REM-D-09-00072.1

Hall, D., J. Scurlock, H. Bolhar-Norden-Kampf, R. Leegood, S. Long, and I. Ross. 1994. Photosynthesis and Production in a Changing Environment; A Field and Laboratory Manual. *Phytochemistry*, 35**:** 278-278.

Hickler, T., L. Eklundh, J. Seaquist, B. Smith, J. Ardö, L. Olsson, M. Sykes, and M. Sjöström. 2005. Precipitation controls Sahel greening trend. *Geophysical Research Letters - GEOPHYS RES LETT*, 32. DOI: 10.1029/2005GL024370

Hilker, T., N. C. Coops, M. A. Wulder, T. A. Black, and R. D. Guy. 2008. The use of remote sensing in light use efficiency based models of gross primary production: A review of current status and future requirements. *Science of the Total Environment*, 404**:** 411-423. DOI: 10.1016/j.scitotenv.2007.11.007

Houghton, R. A., J. House, J. Pongratz, G. Van Der Werf, R. DeFries, M. Hansen, C. Le Quéré, and N. Ramankutty. 2012. Carbon emissions from land use and land-cover change. *Biogeosciences***:** 5125-5142.

Jiang, Z., A. R. Huete, K. Didan, and T. Miura. 2008. Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, 112**:** 3833-3845. DOI: 10.1016/j.rse.2008.06.006

K'Otuto, G. O., D. O. Otieno, B. Seo, H. O. Ogindo, and J. C. Onyango. 2013. Carbon dioxide exchange and biomass productivity of the herbaceous layer of a managed tropical humid savanna ecosystem in western Kenya. 286-297.

Kanniah, K. D., J. Beringer, and L. B. Hutley. 2013. Response of savanna gross primary productivity to interannual variability in rainfall: Results of a remote sensing based light use efficiency model. *Progress in Physical Geography*, 37**:** 642-663. DOI: 10.1177/0309133313490006

Kanniah, K. D., J. Beringer, L. B. Hutley, N. J. Tapper, and X. Zhu. 2009. Evaluation of Collections 4 and 5 of the MODIS Gross Primary Productivity product and algorithm improvement at a tropical savanna site in northern Australia. *Remote Sensing of Environment*, 113**:** 1808-1822. DOI: 10.1016/j.rse.2009.04.013

Kiruki, H. M., E. H. van der Zanden, Ž. Malek, and P. H. Verburg. 2017. Land Cover Change and Woodland Degradation in a Charcoal Producing Semi-Arid Area in Kenya. *Land Degradation & Development*, 28**:** 472-481. DOI: 10.1002/ldr.2545

Kurgat, B. K., E. Ngenoh, H. K. Bett, S. Stöber, S. Mwonga, H. Lotze-Campen, and T. S. Rosenstock. 2018. Drivers of sustainable intensification in Kenyan rural and peri-urban vegetable production. *International Journal of Agricultural Sustainability*, 16**:** 385-398. DOI: 10.1080/14735903.2018.1499842

Lasslop, G., M. Reichstein, D. Papale, A. D. Richardson, A. Arneth, A. Barr, P. Stoy, and G. Wohlfarth. 2010. Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation. *Global Change Biology*, 16**:** 187-208. DOI: 10.1111/j.1365-2486.2009.02041.x

Leuning, R., H. A. Cleugh, S. J. Zegelin, and D. Hughes. 2005. Carbon and water fluxes over a temperate Eucalyptus forest and a tropical wet/dry savanna in Australia: measurements and comparison with MODIS remote sensing estimates. *Agricultural and Forest Meteorology*, 129**:** 151-173. DOI: 10.1016/j.agrformet.2004.12.004

Lloyd, J., and J. A. Taylor. 1994. On the Temperature Dependence of Soil Respiration. *Functional Ecology*, 8**:** 315-323. DOI: 10.2307/2389824

Ma, X., A. Huete, Q. Yu, N. Restrepo-Coupe, J. Beringer, L. B. Hutley, K. D. Kanniah, J. Cleverly, et al. 2014. Parameterization of an ecosystem light-use-efficiency model for predicting savanna GPP using MODIS EVI. *Remote Sensing of Environment*, 154**:** 253-271. DOI: https://doi.org/10.1016/j.rse.2014.08.025

McCree, K. J. 1972. Test of current definitions of photosynthetically active radiation against leaf photosynthesis data. *Agricultural Meteorology*, 10**:** 443-453.

Merbold, L., J. Ardö, A. Arneth, R. J. Scholes, Y. Nouvellon, A. de Grandcourt, S. Archibald, J. M. Bonnefond, et al. 2009. Precipitation as driver of carbon fluxes in 11 African ecosystems. *Biogeosciences*, 6**:** 1027-1041. DOI: 10.5194/bg-6-1027-2009

Monteith, J. L. 1972. Solar Radiation and Productivity in Tropical Ecosystems. *Journal of Applied Ecology*, 9**:** 747-766. DOI: 10.2307/2401901

Oren, R., J. Sperry, G. Katul, D. Pataki, B. Ewers, N. Phillips, and K. Schäfer. 1999. Survey and synthesis of intra- and interspecific variation in stomatal sensitivity to vapour pressure deficit. Plant, Cell & Environment, 22: 1515-1526.

Papale, D., M. Reichstein, M. Aubinet, E. Canfora, C. Bernhofer, W. Kutsch, B. Longdoz, S. Rambal, et al. 2006. Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation. *Biogeosciences*, 3**:** 571-583. DOI: 10.5194/bg-3-571-2006

Poulter, B., D. Frank, P. Ciais, R. B. Myneni, N. Andela, J. Bi, G. Broquet, J. G. Canadell, et al. 2014. Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle. *Nature*, 509**:** 600-603.

Rosenstock, T., M. Mpanda, D. Pelster, K. Butterbach-Bahl, M. Rufino, M. Thiongo, P. Mutuo, S. Abwanda, et al. 2016. Greenhouse gas fluxes from agricultural soils of Kenya and Tanzania: GHG fluxes from ag soils of East Africa. *Journal of Geophysical Research: Biogeosciences*, 121. DOI: 10.1002/2016JG003341

Running, S. W., R. R. Nemani, F. A. Heinsch, M. Zhao, M. Reeves, and H. Hashimoto. 2004. A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production. *BioScience*, 54**:** 547-560. DOI: 10.1641/0006-3568(2004)054[0547:acsmog]2.0.co;2

Sankaran, M., N. P. Hanan, R. J. Scholes, J. Ratnam, D. J. Augustine, B. S. Cade, J. Gignoux, S. I. Higgins, et al. 2005. Determinants of woody cover in African savannas. *Nature*, 438**:** 846-849. DOI: 10.1038/nature04070

Schubert, P., F. Lagergren, M. Aurela, T. Christensen, A. Grelle, M. Heliasz, L. Klemedtsson, A. Lindroth, et al. 2012. Modeling GPP in the Nordic forest landscape with MODIS time series data—Comparison with the MODIS GPP product. *Remote Sensing of Environment*, 126**:** 136-147. DOI: 10.1016/j.rse.2012.08.005

Sigei, J. (2020) Without breaking the soil, farmer turns dry land into grain basket. https://www.nation.co.ke/business/seedsofgold/A-grain-basket-in-the-desert/2301238-5430762-22q4jdz/index.html (last accessed 26 January 2020).

Sjöström, M., J. Ardö, A. Arneth, N. Boulain, B. Cappelaere, L. Eklundh, A. De Grandcourt, W. L. Kutsch, et al. 2011. Exploring the potential of MODIS EVI for modeling gross primary production across African ecosystems. *Remote sensing of environment*, 115**:** 1081-1089.

Sjöström, M., J. Ardö, L. Eklundh, B. A. El-Tahir, H. A. M. El-Khidir, M. Hellström, P. Pilesjö, and J. Seaquist. 2009. Evaluation of satellite based indices for gross primary production estimates in a sparse savanna in the Sudan. *Biogeosciences***:** 129. DOI: 10.5194/bg-6-129-2009

Sjöström, M., M. Zhao, S. Archibald, A. Arneth, B. Cappelaere, U. Falk, A. de Grandcourt, N. Hanan, et al. 2013. Evaluation of MODIS gross primary productivity for Africa using eddy covariance data. *Remote Sensing of Environment***:** 275. DOI: 10.1016/j.rse.2012.12.023

Sombroek, W. G., H. Braun, and B. Van der Pouw. 1982. *Exploratory soil map and agro-climatic zone map of Kenya, 1980. Scale 1: 1,000,000*. Kenya Soil Survey.

Stocker, B. D., J. Zscheischler, T. F. Keenan, I. C. Prentice, S. I. Seneviratne, and J. Peñuelas. 2019. Drought impacts on terrestrial primary production underestimated by satellite monitoring. *Nature Geoscience*, 12**:** 264-270. DOI: 10.1038/s41561-019-0318-6

Tagesson, T., J. Ardö, B. Cappelaere, L. Kergoat, A. Abdi, S. Horion, and R. Fensholt. 2017. Modelling spatial and temporal dynamics of gross primary production in the Sahel from earth-observation-based photosynthetic capacity and quantum efficiency. *Biogeosciences*, 14**:** 1333-1348. DOI: 10.5194/bg-14-1333-2017

Tagesson, T., R. Fensholt, B. Cappelaere, E. Mougin, S. Horion, L. Kergoat, H. Nieto, C. Mbow, et al. 2016. Spatiotemporal variability in carbon exchange fluxes across the Sahel. *Agricultural and Forest Meteorology*, 226-227**:** 108-118. DOI: https://doi.org/10.1016/j.agrformet.2016.05.013

Tagesson, T., R. Fensholt, I. Guiro, M. O. Rasmussen, S. Huber, C. Mbow, M. Garcia, S. Horion, et al. 2015. Ecosystem properties of semiarid savanna grassland in West Africa and its relationship with environmental variability. *Global Change Biology*, 21**:** 250-264. DOI: 10.1111/gcb.12734

Tagesson, T., G. Schurgers, S. Horion, P. Ciais, F. Tian, M. Brandt, A. Ahlström, J.-P. Wigneron, et al. 2020. Recent divergence in the contributions of tropical and boreal forests to the terrestrial carbon sink. *Nature Ecology & Evolution*, 4**:** 202-209. DOI: 10.1038/s41559-019-1090-0

Tan, B., C. E. Woodcock, J. Hu, P. Zhang, M. Ozdogan, D. Huang, W. Yang, Y. Knyazikhin, et al. 2006. The impact of gridding artifacts on the local spatial properties of MODIS data: Implications for validation, compositing, and band-to-band registration across resolutions. *Remote Sensing of Environment*, 105**:** 98-114. DOI: https://doi.org/10.1016/j.rse.2006.06.008

Turner, D. P., W. D. Ritts, W. B. Cohen, T. K. Maeirsperger, S. T. Gower, A. A. Kirschbaum, S. W. Running, M. Zhao, et al. 2005. Site-level evaluation of satellite-based global terrestrial gross primary production and net primary production monitoring. *Global Change Biology*, 11**:** 666-684. DOI: 10.1111/j.1365-2486.2005.00936.x

Wang, L., H. Zhu, A. Lin, L. Zou, W. Qin, and Q. Du. 2017. Evaluation of the latest modis gpp products across multiple biomes using global eddy covariance flux data. *Remote Sensing*, 9**:** 418.

Williams, C. A., N. P. Hanan, I. Baker, G. J. Collatz, J. Berry, and A. S. Denning. 2008. Interannual variability of photosynthesis across Africa and its attribution. *Journal of Geophysical Research: Biogeosciences*, 113. DOI: 10.1029/2008jg000718

Wutzler, T., A. Lucas-Moffat, M. Migliavacca, J. Knauer, K. Sickel, L. Šigut, O. Menzer, and M. Reichstein. 2018. Basic and extensible post-processing of eddy covariance flux data with REddyProc. *Biogeosciences*, 15**:** 5015-5030. DOI: 10.5194/bg-15-5015-2018

Xin, Q., M. Broich, A. E. Suyker, L. Yu, and P. Gong. 2015. Multi-scale evaluation of light use efficiency in MODIS gross primary productivity for croplands in the Midwestern United States. *Agricultural and Forest Meteorology*, 201**:** 111-119. DOI: https://doi.org/10.1016/j.agrformet.2014.11.004

Yengoh, G. T., D. Dent, L. Olsson, A. E. Tengberg, and C. J. Tucker. 2015. Use of the Normalized Difference Vegetation Index (NDVI) to Assess Land Degradation at Multiple Scales: Current Status, Future Trends, and Practical Considerations. 12-13 pp.: Springer International Publishing.

Zhu, Z., S. Piao, R. B. Myneni, M. Huang, Z. Zeng, J. G. Canadell, P. Ciais, S. Sitch, et al. 2016. Greening of the Earth and its drivers. *Nature Climate Change*, 6**:** 791-795. DOI: 10.1038/nclimate3004

Appendix 1

Table S1: The coefficients used in the final two non-linear GPP models based on the Mitcherlich light-response curve following the methodology by Tagesson et al. (2017). The coefficients derived in this study (model A) were found using regression analysis between Sentinel-2 EVI2 and Ausquest farm flux data. The coefficients used in Tagesson et al. (2017) were tested in model B.

Coefficient	**This study (model A)**	**Tagesson et al. (2017) (model B)**
k_fopt	72.6	79.6
m_fopt	-9.5	-7.3
l_fopt	1.91	3.51
n_fopt	0.086	0.03
k_s	0.054	0.16
m_s	0.025	-0.014
l_s	1.836	3.75
n_s	0.067	0.02