Master Thesis in Geographical Information Science nr 169

A comparison of remote sensed semi-arid grassland vegetation anomalies detected using MODIS and Sentinel-3, with anomalies in ground-based eddy covariance flux measurements.

Jason Craig Joubert

2023
Department of
Physical Geography and Ecosystem Science
Centre for Geographical Information Systems
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden

Jason Craig Joubert (2023).

A comparison of remote sensed semi-arid grassland vegetation anomalies detected using MODIS and Sentinel-3, with anomalies in ground-based eddy covariance flux measurements.

Master degree thesis, 30 credits in *Geographical Information Science.*
Department of Physical Geography and Ecosystem Science, Lund University.

Level: Master of Science (MSc).

Disclaimer

This document describes work undertaken as part of a program of study at Lund University. All views and opinions expressed herein remain the sole responsibility of the author, and do not necessarily reflect those of Lund University.

A comparison of remote sensed semi-arid grassland vegetation anomalies detected using MODIS and Sentinel-3, with anomalies in ground-based eddy covariance flux measurements.

Jason Craig Joubert

Master thesis, 30 credits, in Geographical Information Science

Supervisor:

Prof. Jonas Ardö

Dept. of Physical Geography and Ecosystem Science, Lund University

Acknowledgments

I extend my gratitude to my Supervisor, Prof. Jonas Ardö, for his support and guidance during all stages of my study. His knowledge, guidance, input, and motivation were invaluable in supporting the study and the software and systems used herein.

I also extend my thanks to Tim Jacobs from VITO NV for his assistance in providing me with the Python script, and guidance on modifying it for my purposes, that allowed for data sub-setting, spatial focus, and data format conversions. Thanks also to Joshua Gray (NCSU) and Mark Friedl (BU) for granting me permission to reproduce Figure 6 in my work.

I also sincerely thank Lund University, the University of Twente (ITC), and all the educators from these institutions, that tutored me throughout my coursework and further studies.

I also, with immense gratitude, thank my wife, Jenni Joubert, and my twin sons, Tomas, and Dylan, for their patience, understanding, motivation and sacrifices of time lost with me, that allowed me to focus on completing my studies.

Gratitude is also extended to my employer, Apex Emission Testing, for allowing me time between work commitments, to participate in the coursework, and conduct my research and write-up thereof.

Additionally, finally, I thank the Swedish Government, the Italian Government, and the European Union, for the privileges bestowed upon me for being an EU Citizen, that enabled me to take the opportunity to undertake my MSc studies in the first place.

Without the generous contributions and sacrifices of all these individuals, institutions, and entities, neither my studies, nor this thesis, would ever have been possible.

Abstract

Remote sensed vegetation biophysical indicators are derived from various methods and contribute to the quantification of vegetation and crop growth. They are importantly being used as inputs to early warning systems for crop yield estimation, improving crop management and food security.

In this study, the standard score method was used to identify growth anomalies within the remote sensed Gross Primary Production (GPP) of the National Aeronautics and Space Administration's (NASA) Moderate Resolution Imaging Spectroradiometer (MODIS); the Fraction of Absorbed Photosynthetically Active Radiation (*f*APAR); and Gross Dry Matter Productivity (GDMP) observations made by the European Space Agency's PROBA-V and Sentinel-3 satellite-borne multispectral sensors; and compare these against growth anomalies in the ground-based eddy covariance derived GPP biophysical indicator.

The overall aim of the study is to determine the relative strengths of the associations between anomalies detected using these remote sensed indicators, with the ground-based eddy covariance-derived GPP, at a semi-arid grassland site in South Africa.

The findings of the study indicate the strongest ($r = 0.849$, $p < 2.2e\text{-}16$) correlation exists between the anomalies detected using $MODIS_{GPP}$ and those in the reference data (GPP_{EC}), followed by GDMP ($r = 0.667$, $p = 4.3e\text{-}12$), and lastly *f*APAR ($r = 0.239$, $p = 0.045$). Additionally, linear regression analysis revealed that $MODIS_{GPP}$ ($R^2 = 0.748$, RMSE = 0.599 $gC/m^2/day$) appears to more accurately represent changes in GPP_{EC} than GDMP ($R^2 = 0.538$, RMSE = 1.34 $gC/m^2/day$), and *f*APAR ($R^2 = 0.129$, RMSE = 1.88 $gC/m^2/day$).

Based on these findings, and existing literature, it is suggested that *f*APAR requires additional variables to describe GPP, and that differences in performance are likely due to temporal response differences relating to changes in temperature, vapour pressure deficit, soil moisture, light use efficiency and other variables. Additionally, $MODIS_{GPP}$ most closely represents GPP, of the three indicators assessed. The importance of this is that standard scores calculated using ESA's Sentinel-3 GDMP observational data may prove more useful, for semi-arid grassland vegetation anomaly early warning systems, than *f*APAR, and perhaps only slightly less well than NASA's $MODIS_{GPP.}$

Keywords: *ASAP, Anomaly Detection, Agricultural Anomaly, Remote Sensing, fAPAR, GDMP, GPP, MODIS, NASA, Copernicus, Sentinel-3, TerraScope.*

Table of Contents

List of Abbreviations

ASAP	Anomaly hot Spots of Agricultural Production
CGLS	Copernicus Global Land Service
CRS	Coordinate Reference System
Dekad	10-day composite data product
DMP	Dry Matter Productivity
EC	Eddy Covariance
ECMWF	European Centre for Medium-Range Weather Forecasts
EPSG	European Petroleum Survey Group Geodesy
EOS	End of Season
ESA	European Space Agency
***f*APAR**	Fraction of absorbed photosynthetically active radiation
GCOS	Global Climate Observing System
GDMP	Gross Dry Matter Productivity
GJ_T/ha/day	Gigajoules per hectare per day
GPP	Gross Primary Production
GPP_{EC}	GPP calculated from eddy covariance measurements
HDF5	Network Common Data Form
J_P/J_T	Joule PAR per Joule Total Radiation
J_{AP}/J_T	Joule APAR per Joule Total Radiation
Karoo-1	The Karoo-1 Lenient Grazing Site (Study Site) - 31˚25'20.97" S, 25˚1'46.38" E
kgDM/GJ_{AP}	Kilograms of Dry Matter per Gigajoule APAR
kgDM/ha/day	Kilograms of Dry Matter per hectare per day
LCCS	Land Cover Classification System (United Nations)
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NEE	Net Ecosystem Exchange
netCDF	Network Common Data Form
NPP	Net Primary Productivity
NRT	Near Real Time
RS	Remote Sens(ed)(ing)
RT (0-6)	Real Time (cumulative observation designation)
SOS	Start of Season
Study Site	Same as Karoo-1
VM	Virtual Machine
WGS84	World Geodetic System 1984
WMO	World Meteorological Organisation
zGPP_{EC}	z-Scores of GPP_{EC} dataset
z*f*APAR	z-Scores of *f*APAR dataset
zGDMP	z-Scores of GDMP dataset

List of Tables

List of Equations

List of Appendices

List of Figures

Appendix A Figures

Appendix B Figures

1. Introduction

1.1 The importance of semi-arid grasslands

Spatial and temporal variation of vegetation on earth, play vital roles in the global carbon-climate cycle (Martins, Trigo and Freitas, 2020c), with much attention having been focussed on forest ecosystems. Although forests do dominate the mean carbon sink, it has been shown that semi-arid ecosystems dominate the trend and interannual variability, of the global carbon sink (Ahlström *et al.*, 2015). This seasonal and annual variability of ecosystem-atmosphere carbon dioxide fluxes, or net CO_2 exchange, within semi-arid grasslands have been shown to be largely driven by climate variables and land management, such as grazing intensity and water management practices (Rybchak *et al.*, 2020).

Vegetation health also plays other important ecological roles, for example; in the water production cycle by maintaining soil quality, preventing soil erosion and flooding; assisting in the distribution of nutrients; and affecting precipitation rates, temperatures and productivity (Gaberǎčik and Murlis, 2011).

It is often the case that semi-arid landscapes are used as rangelands for grazing (Cadman et al., 2013; Rybchak et al., 2020), and this is true in parts of South Africa, where rangelands occur within the three dominant biomes, which together account for almost 80% of the total land surface of South Africa (Figure 1) (Mucina and Rutherford, 2006; SANBI, 2018).

1.2 Early Warning Systems

Ecological and agricultural studies, relating to vegetation productivity and management, provide tools that can be used for both ecological purposes as well as food security. These tools, techniques, and systems, make use of methods to detect changes within vegetation, as indicated by the biophysical indicators of GPP, *f*APAR, GDMP and others. One such tool, namely "anomaly detection", plays an important role in alerting to potential threats to food security (Peters et al., 2002; Rembold et al., 2015, 2017, 2019), and by monitoring current data, and comparing them to historical patterns for anomalous growth, it is possible to use these in early warning systems for crop management and food security.

1.3 Scientific Problem

Eddy covariance measurement (EC) is the most accurate technique of measuring net CO_2 exchange (Zhang, Zhao and Lin, 2020), and is useful for determining GPP. However, with the existing limited spatial distribution and spatial representation of EC monitoring stations, remote sensed biophysical indicators have the potential to provide widespread inputs to these early warning detection systems.

The performance of any such detection system, however, is affected by the performance of the chosen remote sensed biophysical indicator, e.g., how well the remote sensed indicator represents actual GPP, and therefore the selection of the remote sensed indicator is important.

Such performance based selection, requires evaluation of the performance of various remote sensed indicators, which has previously been undertaken, however, many of these studies focussed on performance evaluation in mid-to-high latitudes in the Northern Hemisphere, such as assessment of $MODIS_{GPP}$ in Northern China (Wang *et al.*, 2013), or Sentinel-3 and MODIS derived *f*APAR, at AmeriFlux stations in North America (Zhang, Zhao and Lin, 2020).

The northern-latitude focussed studies, poses a potential knowledge gap, relating to testing the performance of *f*APAR, GDMP and $MODIS_{GPP}$, in southern hemisphere semi-arid grasslands specifically.

1.4 Aim of the Study

The aim of this study is thus to address this potential knowledge gap, and test the strength of the associations between remote sensed indicators of *f*APAR, GDMP and $MODIS_{GPP}$, with ground-based eddy covariance-derived GPP (GPP_{EC}), at a semi-arid grassland site in South Africa. Consequently, this study aims to add to the understanding of the relative accuracy of, and relationships between, remote-sensed anomalies using *f*APAR, GDMP and $MODIS_{GPP}$, and the higher accuracy ground observed anomalies in eddy covariance measurements.

2. Background

2.1 Biomes of South Africa

As previously mentioned, the South African landscape is dominated by three major biomes (Figure 1), namely the Savanna, Grasslands, and Nama-Karoo biomes, which respectively cover thirty-two-and-a-half percent (32.5%) (Mucina and Rutherford, 2006; Hill, 2013); twenty-seven-point-nine percent (27.9%) (Mucina and Rutherford, 2006; Cadman et al., 2013); and nineteen-and-a-half percent (19.5%) (Mucina and Rutherford, 2006), of South Africa's total land surface area.

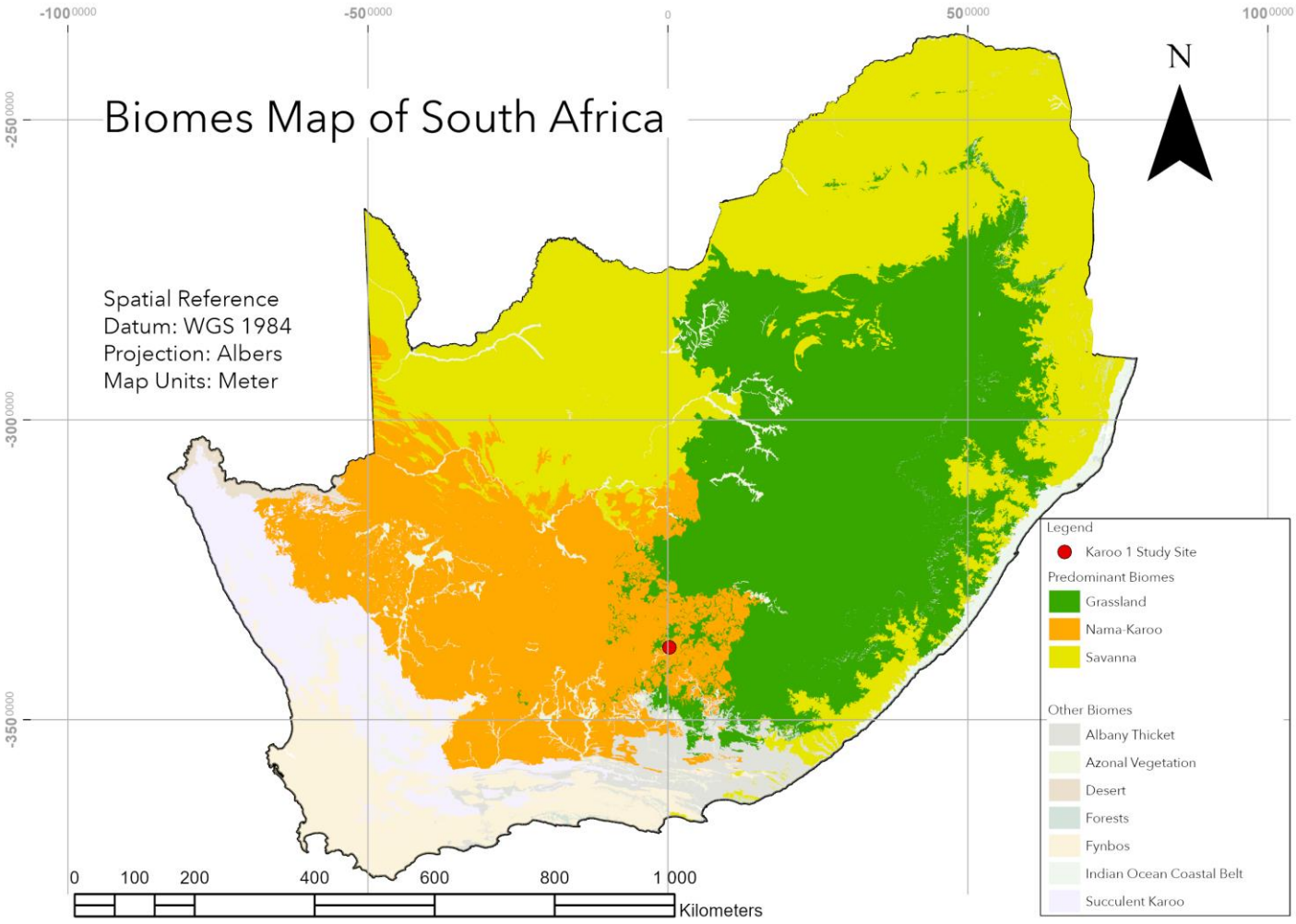


Figure 1 – Biomes Map of South Africa (SANBI, 2018)

2.2 The Nama-Karoo Biome

Of these three biomes, the Nama-Karoo is characterised by being vegetation species-poor with only 14 vegetation units, yet at the same time it holds "many intriguing relationships with its six directly neighbouring biomes" (Mucina and Rutherford, 2006). The boundaries between the Nama-Karoo and its neighbouring biomes are not sharply delineated, e.g., where a distinct change can be seen; but rather range from intermediate transitions over a few kilometres, to

gradual transitions over tens of kilometres (Mucina and Rutherford, 2006). The Nama-Karoo experiences high levels of potential evaporation, and low levels of rainfall, mostly in the late summer and early autumn (Mucina and Rutherford, 2006), with the 1889 to 2023 annual mean rainfall having been reported as 374 mm (range 118 - 731 mm); with a higher annual mean of 445 mm (range 193 - 727 mm), being reported over the past 20 years (du Toit, van den Berg and O'Connor, 2015).

In their classification of South African biomes, Mucina et al. (2006) referenced Heinrich Walter's zonobiomes (Walter 1962, 1968, 1973, 1976, Walter & Box 1946, Walter & Breckle 1991 etc.) and placed the Nama-Karoo in Walter's Arid-Subtropical (zonobiome III), which has characteristics of a desert climate, with very low precipitation; high insolation and light reflection; extreme daily temperature amplitude; and desert to semidesert shrubland vegetation. As is evident from the Savannah being a major neighbouring biome, the transitional boundary of the Nama-Karoo was also identified as falling into Walter's zono-ecotone III-II being "mostly arid plain with isolated low trees (canopy-cover under 10%)" to "climatic savanna on deep, sandy soil with some loam" (Walter and Box, 1976). Walter et. al's (1976) map of zonobiomes and zono-ecotones showed the presence of zonobiome III and zono-ecotone III-II, in other areas in the world, namely, the central to western edges of Australia, South America, India, the Middle East, North Africa, and Mexico.

Despite having a limited number of vegetation types, with 0.5 vegetation units per 1 000 km^2 (Mucina and Rutherford, 2006), the biomass within the Nama-Karoo biome still plays important ecological and economic roles. For example, the areas within the Nama-Karoo have, for decades, been major livestock grazing regions, with a reported 4 million (2007) – 11 million (1939) sheep having been farmed there (Rybchak et al., 2020). The Nama-Karoo is thus an area where complex interactions between humans, flora and fauna, ideally, result in a sustainable cycle of ecological conservation and revenue from agriculture, creating socio, ecological and economic importance to the biome.

Since South African semi-arid ecosystems have this socio-ecological-economic importance, studies within the semi-arid Nama-Karoo biome are important in understanding its' role in the carbon-climate cycle, vegetation, and crop management, as well as food security.

2.3 Study Site

The study site (31°25'20.97"S, 25°1'46.38"E) is located within the Nama-Karoo biome, and specifically at a sheep-grazed grassland area (Figure 3), where an eddy covariance flux tower has been operated since October 2015. This site, termed "Karoo-1" or "The Lenient Grazing Site", was previously the focus of a multi-year eddy covariance study on grazing intensity (Rybchak et al., 2020), and has provided useful Net Ecosystem Exchange (NEE), and GPP data which serves as the reference data for this study. The area around the site (5 km radius) is covered by Grassland (50.8%) and Shrubland (44.5%) (DFFE, 2020), and is at an altitude of 1,310 m.a.s.l (Rybchak *et al.*, 2020), north of the town of Middleburg in the Eastern Cape Province of South Africa. The vegetation, at the site, consists predominantly of grasses and dwarf shrubs, with herbs, sedges and geophytes also being present (Rybchak *et al.*, 2020). The site has a reported temperature range of -4 to 40˚C, and a long term mean annual temperature of 15˚C (Rybchak *et al.*, 2020).

The inter-annual variability within growing seasons, at the study site, is evident within the GPP_{EC} observations at the site (Table 1 and Figure 2). The sum of GPP_{EC} ranged from 16.13 (2016) to 39.52 (2018) $gC/m^2/day$, whilst intra-seasonal standard deviations ranged from 1.06 (2016) to 1.97 (2020), and variances (s^2) ranged from 1.13 (2016) to 3.89 (2020), indicating both inter-annual and intra-seasonal variability at the study site.

Table 1 – Growing Seasons Inter-annual GPP_{EC} at the Study Site ($gC/m^2/day$)

	2016	2017	2018	2019	2020	2021	Range
n	12	14	15	14	13	12	
Minimum	0.13	0.23	0.46	0.02	0.09	0.14	0.44
Maximum	3.03	5.16	5.19	3.69	6.45	5.61	3.42
Sum	16.13	30.27	39.52	21.67	39.04	22.99	23.39
Mean	1.34	2.16	2.63	1.55	3.00	1.92	1.66
Standard Deviation	1.06	1.77	1.61	1.20	1.97	1.97	0.91
Variance (s^2)	1.13	3.12	2.60	1.43	3.89	3.87	2.76

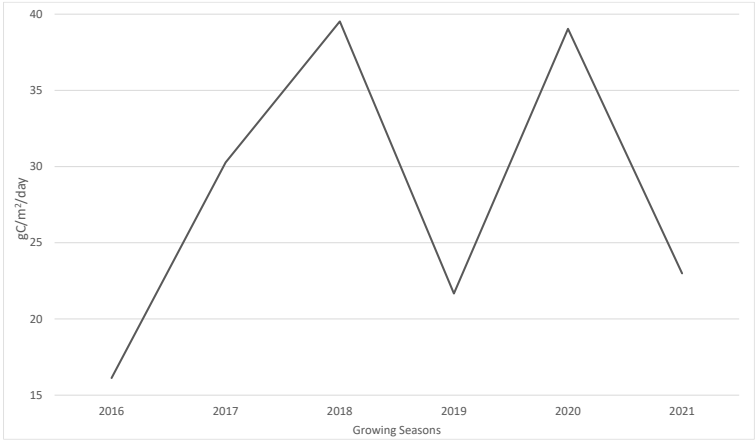


Figure 2 – Plot of seasonal sum of GPP_{EC} showing inter-annual variability ($gC/m^2/day$)

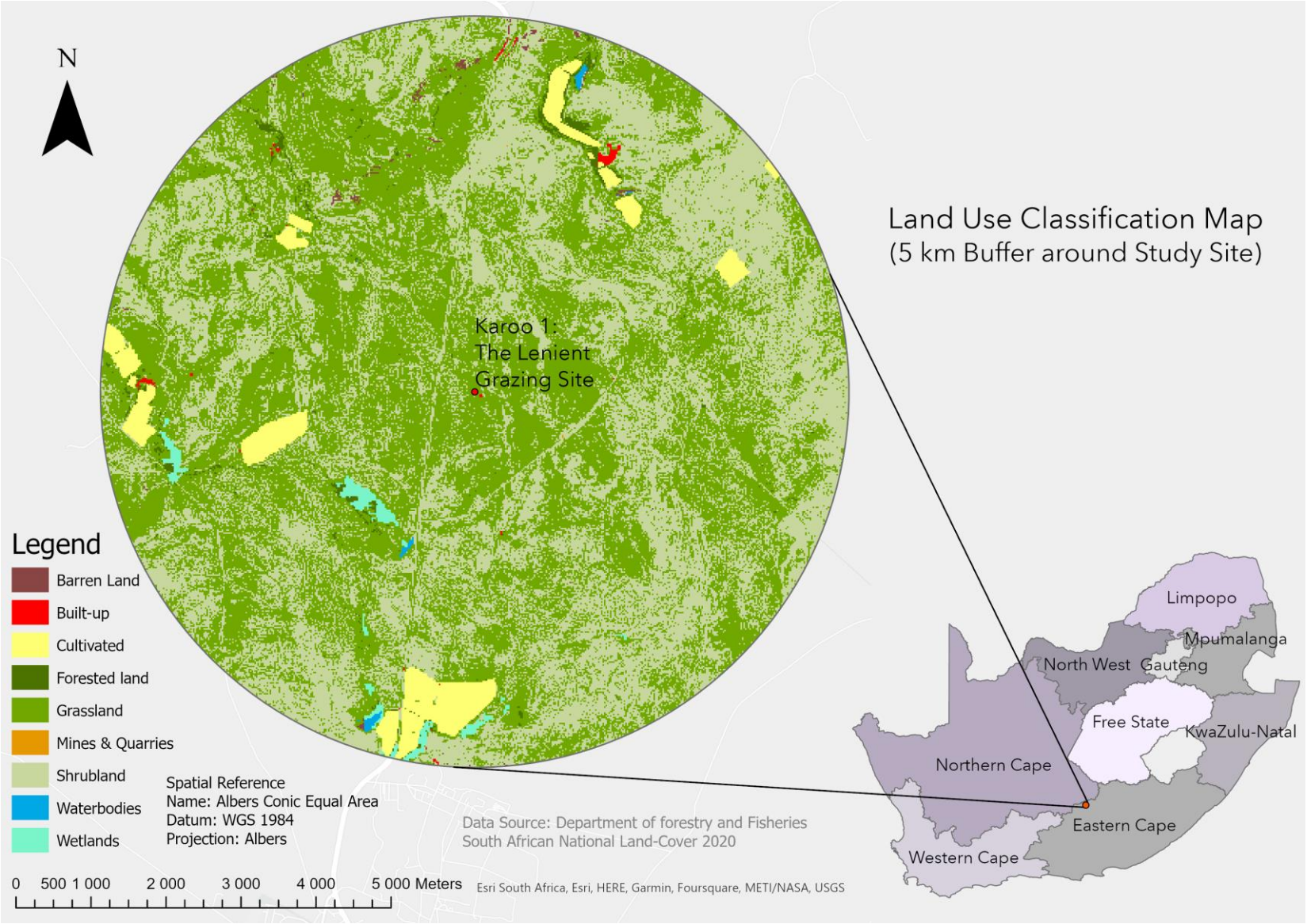


Figure 3 – Land Use Classification Map (showing the location of the Karoo-1 Lenient Grazing, Study Site)

2.4 Research Questions and Hypotheses

Is there a statistically significant association between anomalies detected using:

a. *f*APAR and GDMP?

- (H_a): There is a strong ($r > 0.7$) and statistically significant ($p < 0.01$) association between anomalies detected using *f*APAR and GDMP.

- (H_0): There is not a strong ($r > 0.7$) or statistically significant ($p < 0.01$) association between anomalies detected using *f*APAR and GDMP.

b. *f*APAR and GPP_{EC}?

- (H_a): There is a strong ($r > 0.7$) and statistically significant ($p < 0.01$) association between anomalies detected using *f*APAR and GPP_{EC}.

- (H_0): There is not a strong ($r > 0.7$) or statistically significant ($p < 0.01$) association between anomalies detected using *f*APAR and GPP_{EC}.

c. GDMP and GPP_{EC}?

- (H_a): There is a strong ($r > 0.7$) and statistically significant ($p < 0.01$) association between anomalies detected using GDMP and GPP_{EC}.

- (H_0): There is not a strong ($r > 0.7$) or statistically significant ($p < 0.01$) association between anomalies detected using GDMP and GPP_{EC}.

d. $MODIS_{GPP}$ and *f*APAR?

- (H_a): There is a strong ($r > 0.7$) and statistically significant ($p < 0.01$) association between anaomalies detected using $MODIS_{GPP}$ and *f*APAR.

- (H_0): There is not a strong ($r > 0.7$) or statistically significant ($p < 0.01$) association between anomalies detected using $MODIS_{GPP}$ and *f*APAR.

e. $MODIS_{GPP}$ and GDMP?

- (H_a): There is a strong ($r > 0.7$) and statistically significant ($p < 0.01$) association between anaomalies detected using $MODIS_{GPP}$ and GDMP.

- (H_0): There is not a strong ($r > 0.7$) or statistically significant ($p < 0.01$) association between anomalies detected using $MODIS_{GPP}$ and GDMP.

f. $MODIS_{GPP}$ and GPP_{EC}?

- (H_a): There is a strong ($r > 0.7$) and statistically significant ($p < 0.01$) association between anomalies detected using $MODIS_{GPP}$ and GPP_{EC}.

- (H_0): There is not a strong ($r > 0.7$) or statistically significant ($p < 0.01$) association between anomalies detected using $MODIS_{GPP}$ and GPP_{EC}.

2.5 Hypothesis Test

The Null Hypothesis (H_0) will be rejected, and the Alternative Hypothesis (H_a) will be accepted, if the r value is less than 0.7, and/or, the p-value is below the chosen significance level of 0.01.

3. Methodology

3.1 Anomaly Detection

Anomalies occurring in any series of observational data, can be calculated using several statistical methods. One such method, the standard score method resulting in z-scores, describes the number of standard deviations that an observation is away from the mean of its historical series of observations. The standard score method has been used in studies of vegetation growth anomaly detection and early warning systems (Meroni *et al.*, 2019), and will also be used in this study, where z-scores within remote sensed *f*APAR, GDMP and $MODIS_{GPP}$ data, will be compared against temporally corresponding z-scores of ground observations of GPP, derived using the eddy covariance method.

Since z-scores represent the number of standard deviations that an observation is away from the mean of historical observations, z-scores exceeding defined thresholds can be used to identify anomalies within biophysical indicators. A z-score of zero (0) indicates nominal vegetation growth, whilst z-scores in the positive (+) indicate more abundant vegetation growth, and negative (-) z-scores indicate less abundant vegetation growth, relative to the historical mean of the observed vegetation growth.

The means, and standard deviations, will be calculated for each variable in the datasets, and further used to calculate z-scores, using Equation 1, for each observation. The z-scores will be stored in a set of data termed the *Anomaly Analysis Dataset* (Annex E), comprising of z-scores for each observation, and will be used to assess the strength of associations between *f*APAR, GDMP, $MODIS_{GPP}$ and GPP_{EC}.

$$Z = \frac{x - \mu}{\sigma}$$

Eq. 1.

Where,
Z is the calculated z-score,
x = observed value,
μ = mean of the sample,
σ = standard deviation of the sample.

3.2 Measurement Techniques

3.2.1 Eddy Covariance Flux Measurements

The eddy covariance technique is a micrometeorological method (Baldocchi, Hincks and Meyers, 1988) that can be used for assessing carbon dioxide (CO_2) (and other trace gases), and water and energy exchanges in ecosystems (Baldocchi, 2003; Wutzler et al., 2018).

The technique functions by applying a conservation equation on the assumption that a rate of change of a gas' mixing ratio, e.g. a mass of CO_2 per unit mass of dry air (at a fixed point), can be determined from turbulent fluxes (derived from the mean covariances between mixing ratio and the mean horizontal and vertical wind velocities); molecular diffusion; and a-priori knowledge of whether the measurement is for a source, or a sink (Baldocchi, Hincks and Meyers, 1988).

Eddy covariance flux data yields results of net ecosystem exchange (NEE), or net CO_2 balance ($gC/m^2/day$) (Pinker *et al.*, 2010), and its predicted components of GPP (GPP_{EC}) and Ecosystem Respiration (R_{eco}). This is achieved through the application of a chosen flux-partitioning algorithm, e.g. the night-time approach (Reichstein *et al.*, 2005), or the day-time approach (Lasslop *et al.*, 2010), or others (Wutzler *et al.*, 2018).

The differences between the night-time and day-time approaches, briefly, is that Reichstein *et al's* (2005) night-time approach assumes GPP to be zero at night and therefore any NEE that is measured is assumed to be entirely $R_{eco.}$ This allows for a model to then be fit, that extrapolates night-time data to day-time periods (Lasslop *et al.*, 2010).

The day-time approach on the other hand offered an improved method that extrapolated respiration (R_{eco}) from light-response curves that considered the effects of vapour pressure deficits, and temperature sensitivity of respiration (Lasslop *et al.*, 2010). Simply put however, whichever approach is used, GPP_{EC} and R_{eco} can be estimated according to equation 2 (Reichstein *et al.*, 2005).

$$NEE = R_{eco} - GPP_{EC} \qquad \textbf{Eq. 2.}$$

3.3 Biophysical Indicators

3.3.1 Fraction of Absorbed Photosynthetically Active Radiation

Monteith (1972) proposed that green leaves require solar radiation in the 0.4 – 0.7 μm waveband for photosynthesis, and called this "photosynthetically active radiation" (PAR). Monteith described how the amount of radiation within this waveband was affected by absorption and scattering by dust content and water vapour within the atmosphere, as well as diffuse radiation scattered by gas molecules, and that Moon (1940) had concluded that PAR was close to 50% of total incoming solar radiation (Monteith, 1972).

Monteith (1972) also stated that the fraction of PAR absorbed by green vegetation (*f*APAR) depended on the amount of chlorophyll per unit area of the green vegetation, and proposed that fraction to be approximately 0.85, midway between 80-90% of PAR. The amount of PAR absorbed by green vegetation (APAR) is a function of PAR and *f*APAR and can be expressed as Equation 3 (Ardö *et al.*, 2022).

$$\mathrm{APAR} = \mathrm{PAR} \cdot fAPAR \qquad \textbf{Eq. 3.}$$

*f*APAR is commonly estimated using satellite data, either as a function of the Normalised Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Enhanced Vegetation Index 2 (EVI2), or other methodology (Ardö *et al.*, 2022). This remote sensed *f*APAR "...quantifies the photosynthetic capacity of green vegetation..." (Bacour *et al.*, 2006) and is expressed as a unitless fraction of the incoming radiation received at the land surface. It corresponds to direct sun illuminated, or black-sky only, absorption by green vegetation, which excludes diffuse radiation (Bacour *et al.*, 2006; Baret *et al.*, 2007).

*f*APAR can be used as a standalone product in vegetation studies (Meroni *et al.*, 2014; Durgun *et al.*, 2016; De Lemos, Verstraete and Scholes, 2020), and can also be used in combination with European Centre for Medium-Range Weather Forecasts (ECMWF) modelled meteorological data; biome specific land cover information; and the UN Land Cover Classification System (LCCS); to generate other important vegetation data products, namely, GDMP (Equation 4.a) (Martins, Trigo and Freitas, 2020b, 2020a), as discussed in Section 3.3.3.

3.3.2 Gross Primary Production

GPP refers to the rate at which vegetation produces organic compounds through photosynthesis, represented as grams of carbon per square metre per day ($gC/m^2/day$), and involves the mechanisms of carbon dioxide (CO_2) uptake, release of oxygen (O_2), and the growth and accumulation of biomass. GPP can be estimated using various techniques, models, and methods, both ground-based and remote sensed. GPP data to be used in this study, include remote-sensed data from the MODIS, and ground-based EC flux tower measurements.

3.3.3 Gross Dry Matter Productivity

GDMP is comparable with GPP, except that GDMP is expressed in units of kilograms of dry matter per hectare per day (kgDM/ha/day), since this is a better suited unit for agro-statistical purposes (Durgun *et al.*, 2016). GDMP represents the total amount of dry matter per time unit from photosynthesis, with a significant fraction of GDMP supporting autotrophic respiration (R_a) (Martins, Trigo and Freitas, 2020b).

As discussed previously, GDMP is derived from various input variables including *f*APAR, where according to Martins *et al.* (2020b), GDMP is derived by simplifying equation 4a (and dropping the factor ε_{RES}), to where GDMP can be determined by the four basic variables of *f*APAR, radiation (R), temperature (T) and CO_2, expressed as Equations 4b & 4c.

This method not only simplifies the equation, but also importantly allows for a method that bypasses the spatial and temporal differences in the inputs (Martins, Trigo and Freitas, 2020b, 2020a).

$$\mathrm{GDMP} = \mathrm{R} \cdot \varepsilon_c \cdot f\mathrm{APAR} \cdot \varepsilon\mathrm{LUE}_c \cdot \varepsilon_T \cdot \varepsilon\mathrm{CO}_2 \, [\cdot \varepsilon_{RES}] \qquad \textbf{Eq. 4.a.}$$

Where:
R = total shortwave radiation (0.2 – 3.0 µm) (Value Range 0 – 320 $GJ_T/ha/day$),
ε_c = fraction of PAR (0.4 – 0.7 µm) in total shortwave (Value 0.48 J_P/J_T),
*f*APAR = PAR-fraction absorbed (PA) by green vegetation (Value Range 0.0 … 1.0 J_{AP}/J_P),
εLUE_c = biome specific light use efficiency at optimum (Value Range Biome - specific $kgDM/GJ_{AP}$),
ε_T = normalised temperature effect,
εCO_2 = normalised CO_2 fertilisation effect,
ε_{RES} = fraction kept after omitted effects (drought, pests etc.)

$$\mathrm{GDMP}_{\mathrm{max}} = \mathrm{R} \cdot \varepsilon_c \cdot \varepsilon_T \cdot \varepsilon_{co2} \qquad \textbf{Eq. 4.b.}$$

$$\mathrm{GDMP} = fAPAR \cdot \varepsilon\mathrm{LUE}_c \cdot GDMP_{max} \qquad \textbf{Eq. 4.c.}$$

3.4 Data Access & Processing

3.4.1 Virtual Research Environment

ESA's remote sensed Sentinel-3 *f*APAR and GDMP data were accessed and pre-processed on a TerraScope hosted virtual machine (VM) (VITO NV, 2022), using Python and Linux Bash Scripts (Appendices E - F). The Python script to access and process the netCDF format *f*APAR & GDMP data (Jacobs and VITO NV, 2020) made use of utilities in the open-source "Geospatial Data Abstraction Library" (GDAL/OGR Contributors, 2020), to spatially subset the netCDF global datasets, into 3-x-3-pixel matrices, and convert this data into appropriate formats for this study.

The 3-x-3-pixel matrix was chosen over a single pixel, because flux measurements do not represent a measurement at a single point e.g., where the EC flux tower is located, but rather represent the average of measurements of the surrounding area, the "flux footprint" (Kljun *et al.*, 2015). The spatial orientation and extent of the flux footprint, is dynamic and depends on tower design, surrounding surface roughness, and prevailing environmental conditions at the time of measurement. Therefore, it was decided that an average of the site pixel, and adjacent pixels, would provide better representation than a single pixel alone. Additionally, a larger matrix was not used, simply based on Tobler's First Law of Geography, being "…everything is related to everything else, but near things are more related than distant things." (Tobler, 1970).

Figure 4 displays an example of the image for the study site on the 10/04/2020, displaying the individual pixel *f*APAR values, after applying the Scale Factor of 1/250, or 0.004, (Martins, Trigo and Freitas, 2020c) required to convert DN to scaled *f*APAR values. The average of the example *f*APAR values is 0.454, corresponding to the red bold text observation in Appendix D (Record Number 231).

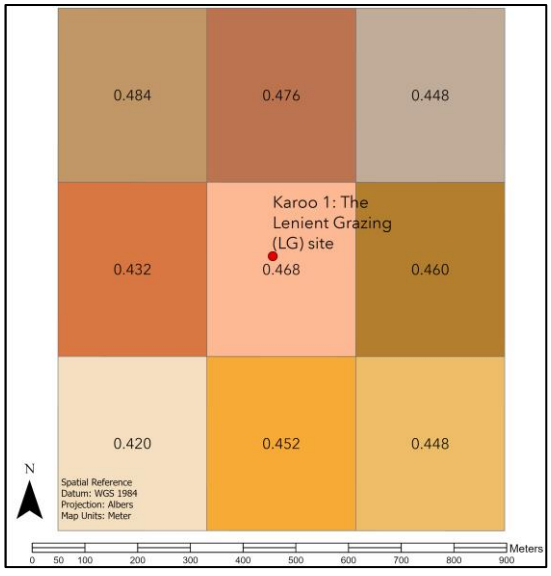


Figure 4 – Example of 3-x-3-pixel showing fAPAR values

3.4.2 Data Flags

The *f*APAR and GDMP data products are provided in Near Real Time (NRT) estimates from real time (RT0) to a 60-day composited (RT6) value. The value of the product at real time (RT0) is considered unstable, relative to subsequent adjustments e.g., RT2, RT5 and RT6, especially when considering noisy and missing data. The process of using less noisy, and more observations as time progresses, improves the accuracy of the product value (Martins, Trigo and Freitas, 2020c), in simple terms, RT6 is considered less noisy, and more representative, than products with lower RT values.

However, it is still noteworthy that a published error evaluation of the consistency of RT0 with RT6 for *f*APAR data products revealed *f*APAR (RT0) to have a 2.1% positive bias (R=0.99), and showed significant improvement after RT1; with stability being achieved after RT2 (RT2 to RT6: R=1.00, RMSD = 0.013) (Martins, Trigo and Freitas, 2020c). Similarly, a Quality Assessment of the differences between RT0, RT1, RT2 and RT6, of the GDMP data product, were found to be "small" (Martins, Trigo and Freitas, 2020b).

3.4.3 *f*APAR and GDMP Data Compositing

The different RT flagged datasets, as downloaded, comprised of large amounts of temporally variable, and temporally disparate, data e.g., some data was missing for some dekads in dataset RT*x*, whilst being present in RT*y*, and vice versa. The disparity across the time coverages of the different datasets, is evident in the graphic depiction (Figure 5), which displays the comparative temporal availability of each dataset. This disparity necessitated ignoring the lesser RT flagged data where both were present, and composting different RT designations for *f*APAR and GDMP, namely RT6 appended to RT2 for *f*APAR, and RT6 appended to RT5 for GDMP, as illustrated by the red bounding boxes in Figure 5. (See Addendum A)

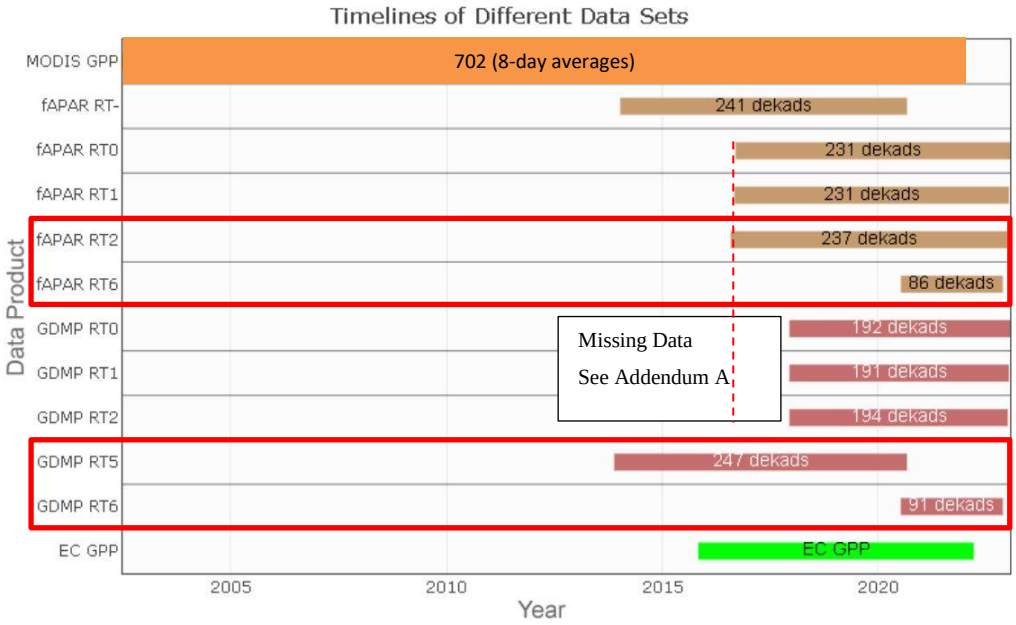


Figure 5 – Comparison of Timelines of Various Datasets

3.4.4 Phenophase Sub-Setting Process

Lieth (1973) described phenology generally as "… the art of observing life cycle phases or activities of plants and animals in their temporal occurrence throughout the year." (Lieth, 1973). In biomass studies these phases have previously been described as "Leaf-out – Growth – Maturity – Stress – Dormancy" (Haynes *et al.*, 2019); broadly as "Growth Phase – Senescence Phase" (De Lemos, Verstraete and Scholes, 2020); and more descriptively as "Greenup – MidGreenup – Maturity – Peak – Senescence – MidGreenDown – Dormancy" (Sulla-Menashe and Friedl, 2018); amongst other descriptions (Richardson *et al.*, 2013; Tian *et al.*, 2021). “Biomass phenology has a clear ecological interpretation and is the key variable of interest for potential users interested in raising livestock or maintaining wildlife.” (https://www.ramona.earth, 2022).

Additionally, accurate representation of phenology in models concerned with land surface and climate systems is of particular importance (Richardson *et al.*, 2013). Hence, knowledge of the phenological stages, within this study, is important because the different vegetation development stages between “Greenup” and “Dormancy”, provide temporal boundaries that have been used to subset the study data, on which baselines were determined, and standard scores were calculated. In other words, anomalies were only identified within the phenophases from “Greenup” to “Dormancy”, similar to the approach adopted in the "Anomaly hot Spots of Agricultural Production" (ASAP) early warning system (Meroni *et al.*, 2019). MODIS data provides phenological information for its data products, from which it was possible to extract the "Start of Season" (SOS) and "End of Season" (EOS) information as summarised in Table 2.

For this study, "Start of Season" (SOS) was taken to be "Greenup", and "End of Season" (EOS) was taken to be "Dormancy", as described in the *User Guide to Collection 6.1 MODIS Land Cover Dynamics* (Sulla-Menashe and Friedl, 2018), and graphically depicted in the example extracted, and appearing here as Figure 6.

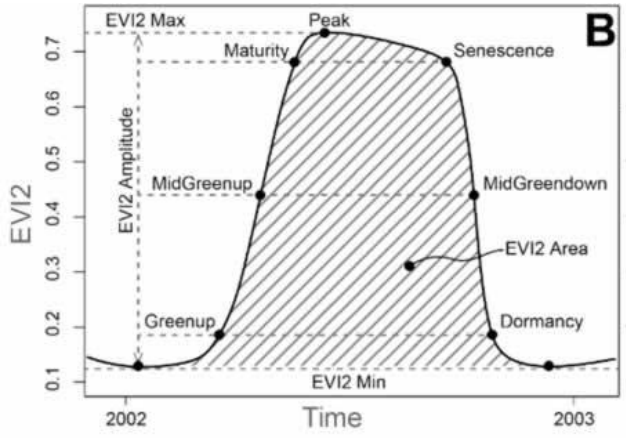


Figure 6: Illustration of Greenup & Dormancy.
Reproduced with permission: (Sulla-Menashe and Friedl, 2018)

Table 2 - Greenup and Dormancy Dates Used to Define the Study Periods

Year	Greenup		Dormancy	
	Day	Date	Day	Date
2014	299	26 Oct '13	149	29 May '14
2015	310	06 Nov '14	169	18 Jun '15
2016	8	08 Jan '15	136	15 May '16
2017	358	24 Dec '16	136	16 May '17
2018	13	13 Jan '18	167	16 Jun '18
2019	15	15 Jan '19	151	31 May '19
2020	1	01 Jan '20	139	18 May '20
2021	352	18 Dec '20	112	22 Apr '21

3.5 Datasets

3.5.1 The MODIS Long-Term (Full) Dataset

NASA's MODIS instruments conduct Earth observations in thirty-six (36) spectral bands from 620 – 14385 nm, providing data products that are used for lower atmosphere, land and ocean study purposes (NASA, 2023). "MYD17A2HGF Version 6.1" is a gap-filled data product and includes long-term (2002 – Present) 500 metre spatial resolution, 8-day total GPP observations in $kgC/m^2/8d$ (Running, S., Mu, Q., Zhao, 2021), which was converted to $gC/m^2/day$ prior to analysis.

A subset of this dataset, beginning at 20/11/2013, was used to compile a baseline dataset, herein called *"The MODIS Long-Term (Full) Dataset*", which provided an overview of the long-term remote-sensed GPP trends at the study site, and was used to guide temporal parameters used in this study, by using it to determine the phenological cycles, and identify "phenophases" (National Phenology Network (USA-NPN), 2023) of interest for the study location.

These phenophases were subsequently used to create a subset dataset, herein termed *"The MODIS Phenophases-Subset Dataset"*, which was then used to calculate z-scores; conduct correlation analysis with (z)*f*APAR, (z)GDMP and (z)GPP_{EC}; and lastly used for linear regression with the ground-based reference EC observations (GPP_{EC}).

The "phenophase sub-setting process" described previously, used "Start of Season" (SOS), and "End of Season" (EOS) dates (Table 2), which were also used to subset the composited *f*APAR, GDMP, $MODIS_{GPP}$ and GPP_{EC} datasets.

"The MODIS Long-Term (Full) Dataset" has been graphically depicted in Figures 7, 8 and 9, and *"The MODIS Phenophase-Subset Dataset"* has been depicted in Figures 12 and 14.

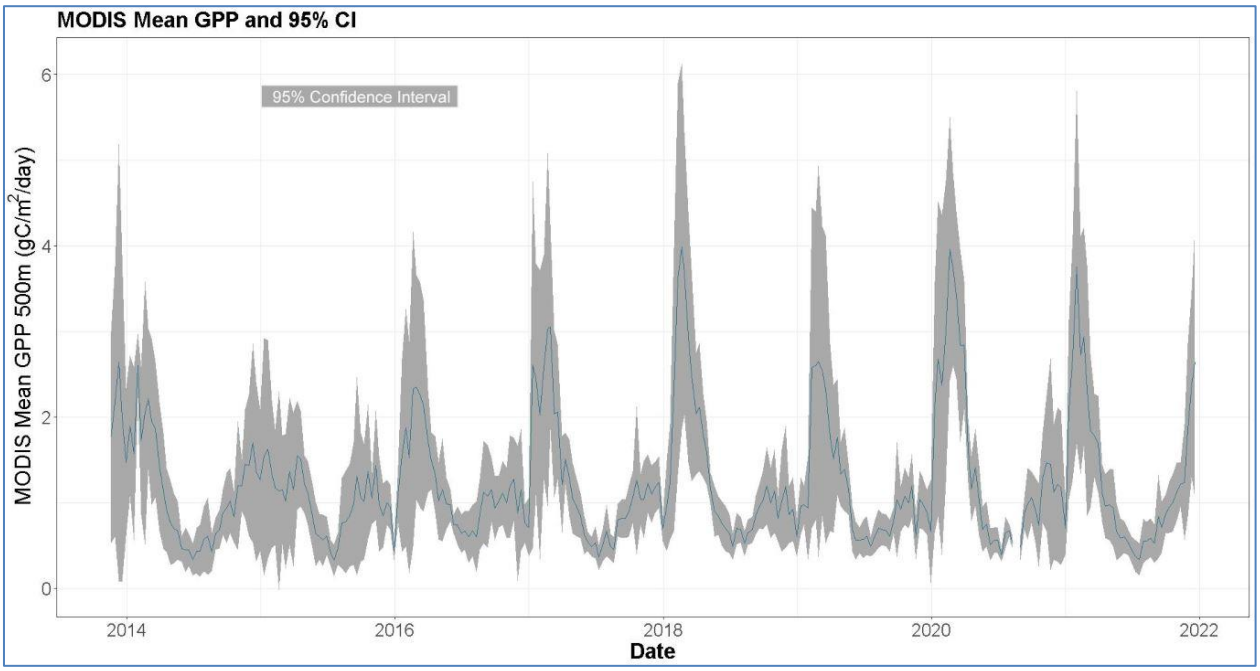


Figure 7 – Timeseries of MODIS Mean GPP (gC/m2/day)
(Single 500m x 500m pixel at Karoo-1)

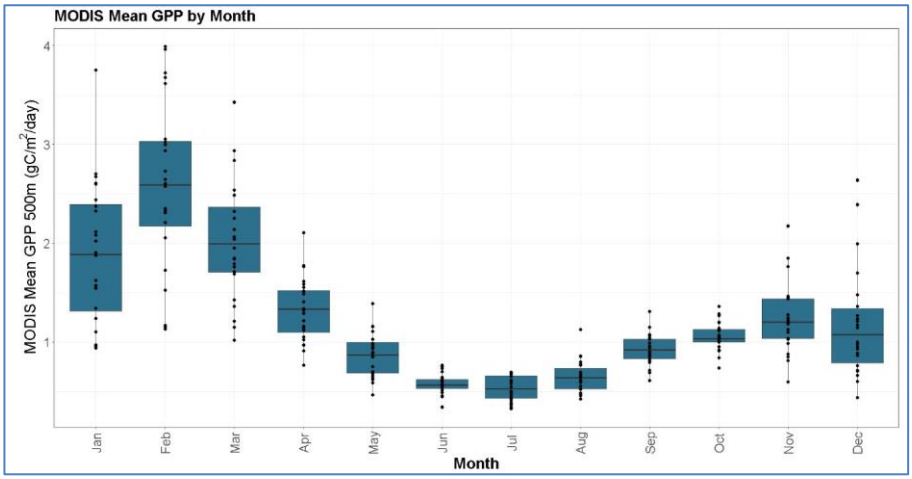


Figure 8– Box Plot of MODIS Mean GPP by Month (2013 – 2021)
(Single 500m x 500m pixel at Karoo 1)

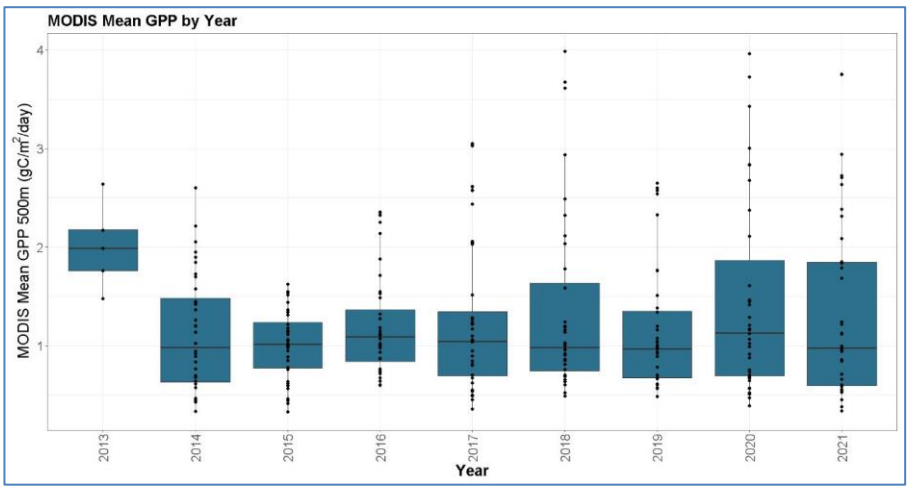


Figure 9 – Box Plot of MODIS Mean GPP by Year (2013 – 2021)
(Single 500m x 500m pixel at Karoo 1)

3.5.2 Eddy Covariance Flux Station Data

For this study, the GPP_{EC} data (Rybchak *et al.*, 2020) was derived using the day-time approach of Lasslop et al. (2010), and is the flux component of interest in this study, since it has been used as reference data. The data provided was in micromoles/m^2/sec and was converted to gC/m^2/day prior to analysis. The GPP_{EC} data were provided in a single dataset starting from 10/11/2015 to 10/03/2022 (Thünen Institute, 2022; Thünen Institute of Climate-Smart Agriculture, 2022), in micromoles/m^2/second, 30-minute average observations, and were first converted to gC/m^2/day, then aggregated to daily averages, and finally composited to 10-day averages to temporally match the remote sensed dekad datasets. No information regarding gap-filling of this dataset was able to be ascertained, and the data was taken to be representative, without unknown gap filling procedures, or pre-processing, other than as described here.

3.5.3 *f*APAR Data

*f*APAR is monitored globally by both terrestrial and satellite-borne sensors (World Meteorological Organization (WMO), 2016). *f*APAR is useful as an input variable to primary productivity models (Martins, Trigo and Freitas, 2020c), and is another of the remote sensed biophysical indicators tested in this study. The Copernicus Global Land Service (CGLS) *f*APAR data product is one of the Essential Climate Variables (ECV's) identified by the Global Climate Observing System (GCOS) operated by the WMO. The European Space Agency (ESA) has been remotely sensing *f*APAR, using various sensors, for many years, importantly (for this study), by using the PROBA-V (January 2014 to June 2020), and more recently the Sentinel-3 Ocean and Land Colour Instrument (OLCI) (July 2020 to current), under the CGLS (Martins, Trigo and Freitas, 2020c). The CGLS's "Fraction of Absorbed Photosynthetically Active Radiation Collection 300m" (Version 1.0 & 1.1) data product, comprise of 10-day composites, or "dekads", at 300 metre spatial resolution, and is projected as a regular grid (latitude & longitude) in the plate carrée equidistant cylindrical map projection, with a 1/336° resolution (Martins, Trigo and Freitas, 2020c).

It was observed that a dataset with no RT designation was present for *f*APAR, labelled as "RT-" (Figure 5). This dataset was initially thought to be a good candidate for use, since it had the earliest start date of the *f*APAR data, however, exploratory correlation analysis of corresponding dekads (10/01/2017 to 31/12/2019) between RT- and RT2 showed a weak correlation (R = 0.439), and considering the small differences between *f*APAR RT2 & RT6 previously described (Martins, Trigo and Freitas, 2020c), it was decided that the final analysis be conducted on a *f*APAR composited dataset, comprising of the earliest RT2 (20/08/2016) to the dekad before the earliest RT6 (31/08/2020), followed by the earliest RT6 (10/09/2020), to the end of RT6 (31/12/2022). In other words, RT2 data was used to the date where RT6 data became available, and RT6 was then appended to the RT2 data from that date on. The complete *f*APAR dataset available, for comparison with GPP_{EC}, thus spanned from 20/08/2016 to 31/12/2022 (Table 3).

3.5.4 GDMP Data

The CGLS's Dry Matter Productivity Collection 300m (Version 1.0 & 1.1) data products represent the daily growth of standing biomass. Version 1.0 (January 2014 – June 2020) was derived from *f*APAR using PROBA-V data, and Version 1.1 (July 2020 onwards) was derived from *f*APAR using Sentinel-3/OLCI sensor data. This dataset contains two components, namely Gross Dry Matter Productivity (GDMP) and Dry Matter Productivity (DMP), comprising of 10-day composites (dekads), with 3 dekads per month e.g., days 1 to 10, days 11 to 20, and days 21 to the last day of the month (Martins, Trigo and Freitas, 2020b). The products are also projected as a regular grid (latitude & longitude) in the plate carrée equidistant cylindrical map projection, with a 1/336˚ resolution.

Preliminary assessment of the GDMP datasets revealed limited availability of RT6, starting from 10/09/2020 to 31/12/2022. The data also revealed RT5 availability from 20/11/2013 to 30/09/2020, overlapping RT6 by a few dekads. It was thus decided to compile a composited dataset comprising of RT5 (20/11/2013 – 31/08/2020) followed by RT6 (10/09/2020 to 10/12/2022), in a similar method applied to *f*APAR, described previously. The complete GDMP dataset available, for comparison with GPP_{EC}, thus spanned from 20/11/2013 to 10/12/2022 (Table 3).

Table 3 – Table of Individual Study Periods and Combined Study Periods

Biophysical Indicator	Individual Study Periods		Combined Composited Periods (No. of Dekads)
	Start	End	
***f*APAR (RT2)**	20/08/2016	31/08/2020	20/08/2016 - 31/12/2022 (189)
***f*APAR (RT6)**	10/09/2020	31/12/2022	
GDMP (RT5)	20/11/2013	31/08/2020	20/11/2013 –10/12/2022 (284)
GDMP (RT6)	10/09/2020	10/12/2022	
$MODIS_{GPP}$	20/11/2013	20/12/2021	20/11/2013 – 20/12/2021 (291)
GPP_{EC}	10/11/2015	20/03/2022	10/11/2015 - 20/03/2022 (221)

3.6 The Analyses Datasets

The Methodological Flowchart (Figure 10), includes the data processing and analyses steps that were necessary to remove temporal disparity between datasets; to aggregate the datasets to the common time averages; to calculate z-scores; and conduct statistical analysis. To explore and describe the data, and associations between different data sources, analysis was conducted on three analysis datasets, namely the *"The Full Dataset"- (Annex C)*, "*The Interim Analysis Dataset (DN's)" - (Annex D)*, and the *"Anomaly Analysis Dataset (z-scores)" - (Annex E)*.

3.6.1 Description of "The Interim Analysis Dataset"

The "*Interim Analysis Dataset"* (Annex D) comprised of temporally corresponding observations between the earliest GDMP observation (20/11/2013) and the latest observation within the downloaded datasets (20/12/2021), considering Greenup and Dormancy dates (Table 2). This *Interim Analysis Dataset* comprised of *f*APAR (n = 71 dekads), GDMP (n = 127 dekads), GPP_{EC} (n = 84 dekads) and $MODIS_{GPP}$ (n = 127, 8-day averages). Zero value observations were converted to "NA's" and excluded from the analyses.

3.6.2 Description of "The Anomaly Analysis Dataset (z-scores)"

The "*Anomaly Analysis Dataset (z-scores)"* - (Annex E), comprised of the z-scores derived using equation 1, for each observation, and were used to assess the strength of associations between (z)*f*APAR, (z)GDMP, (z)$MODIS_{GPP}$, and (z)GPP_{EC}.

3.7 Methodological Flowchart

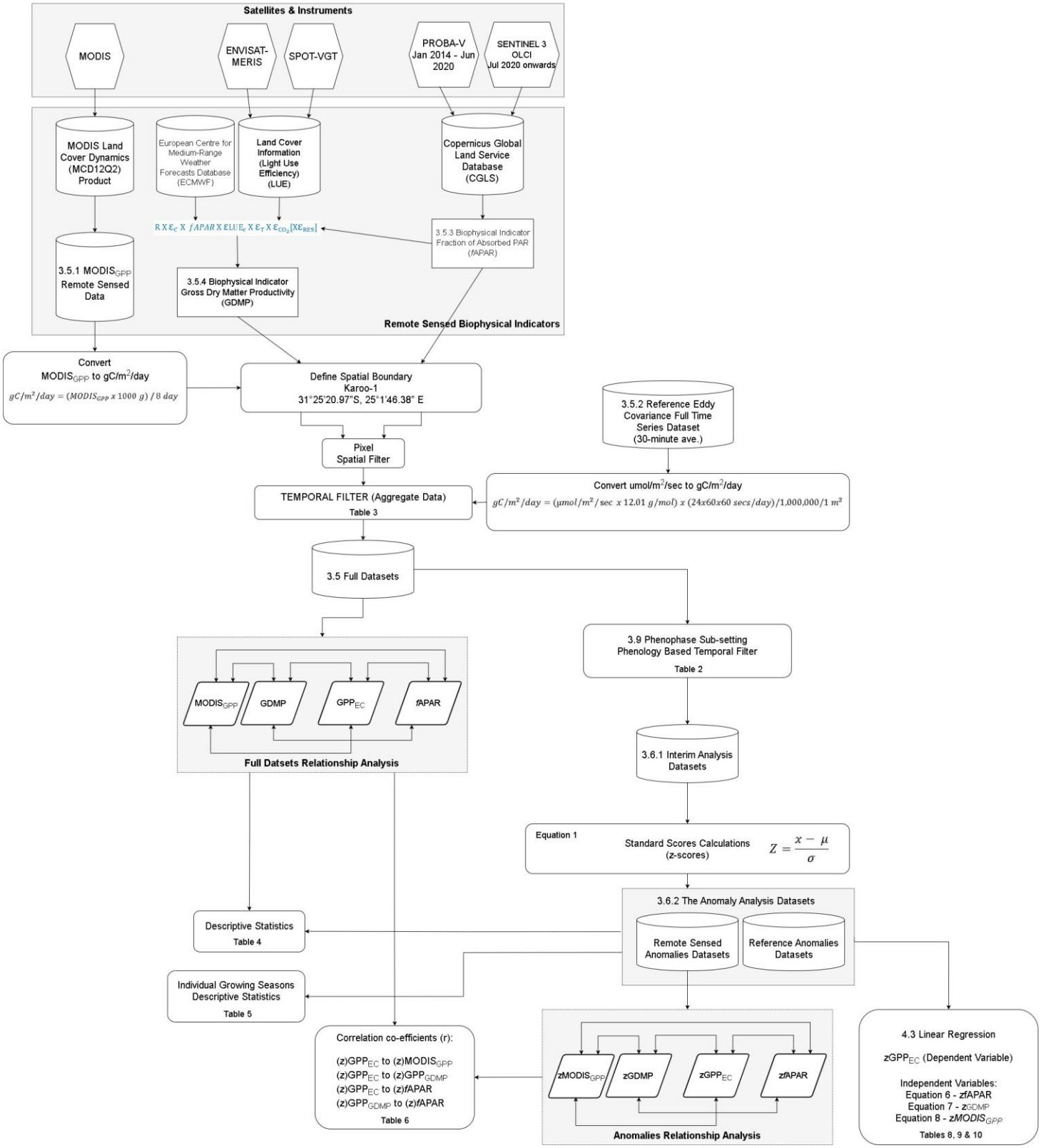


Figure 10 – Methodological Flowchart
(Numbering, equations and tables correspond to those in this document)

4. Results

4.1 Descriptive Statistics

Exploratory analysis (Annex H) was conducted on the dataset using the R Language and Environment for Statistical Computing (R Core Team, 2021). Descriptive statistics for the various datasets have been summarised in Table 4.

Table 4 – Descriptive Statistics of the Full, Interim and Anomaly Analysis Datasets

Full Datasets (Including All Phenological Phases)								
Variable	Min	1st Quarter	Median	Mean	3rd Quarter	Max.	S. D	n
*f*APAR (Unitless)*	0.04	0.06	0.09	0.15	0.21	0.54	0.13	189
GDMP (kgDM/ha/day)	5.1	620.8	1005.7	1669.1	2136.4	8039.6	1665.2	284
GPP_{EC} ($gC/m^2/day$)	0.0032	0.0615	0.2671	0.9233	1.0017	6.4536	1.3862	221
$MODIS_{GPP}$ ($gC/m^{2/}day$) #	0.325	0.6937	1.0250	1.2426	1.5125	3.9875	0.7445	291

* *f*APAR Unitless = *f*APAR(DN) x (1/250) (Martins, Trigo and Freitas, 2020c)

$MODIS_{GPP}$ unit conversion - convert kg to g and divide by 8 days (Running et al., 2019)

Interim Analysis Dataset (Greenup-to-Dormancy Phenophase Subsets)								
Variable	Min	1st Quarter	Median	Mean	3rd Quarter	Max.	S. D	n
*f*APAR (Unitless)*	0.04	0.09	0.26	0.26	0.36	0.54	0.15	71
GDMP (kgDM/ha/day)	208.1	1318.2	2338.2	2799.2	3748.1	8039.6	1878.4	127
GPP_{EC} ($gC/m^2/day$)	0.019	0.634	1.479	2.040	3.347	6.454	1.681	84
$MODIS_{GPP}$ ($gC/m^{2/}day$) #	0.463	1.150	1.575	1.758	2.319	3.988	0.808	127

* *f*APAR Unitless = *f*APAR(DN) x (1/250) (Martins, Trigo and Freitas, 2020c)

$MODIS_{GPP}$ unit conversion - convert kg to g and divide by 8 days (Running et al., 2019)

Anomaly (z-scores) Analysis Dataset								
Variable	Min	1st Quarter	Median	Mean	3rd Quarter	Max.	S. D	n
z*f*APAR (z-score)	-1.417	-1.138	0.031	0.000	0.658	1.905	1.000	71
zGDMP (z-score)	-1.379	-0.789	-0.245	0.000	0.505	2.789	1.000	127
$zGPP_{EC}$ (z-score)	-1.202	-0.833	-0.333	0.000	0.777	2.625	1.000	84
$zMODIS_{GPP}$ (z-score)	-1.603	-0.753	-0.227	0.000	0.693	2.758	1.000	127

The Anomaly Analysis Dataset (z-scores) was further explored (Figure 11), at individual growing season levels, to seek an understanding of the role that each of the remote sensed biophysical indicators could play in vegetation growth anomaly detection, and potentially as early warnings of a failure within growing seasons. Descriptive statistics of each of the individual years have been summarised in Table 5.

Table 5 – Descriptive Statistics of the Anomaly Analysis Dataset (z-scores) of the Individual Growing Seasons

***f*APAR (z-scores)**							
Year	Min	Median	Mean*	Max	S.D	Sum	n
2014	-	-	-	-	-	-	0
2015	-	-	-	-	-	-	0
2016	-	-	-	-	-	-	0
2017	-1.367	-0.046	**-0.013**	1.529	0.894	-0.179	14
2018	-1.417	0.643	**0.356**	1.819	1.038	5.335	15
2019	-1.325	-0.097	**-0.187**	0.747	0.806	-2.622	14
2020	-1.405	0.948	**0.608**	1.905	1.088	7.909	13
2021	-1.331	-1.209	**-0.686**	0.635	0.788	-8.234	12
GDMP (z-scores)							
Year	Min	Median	Mean*	Max	S.D	Sum	n
2014	-1.160	-0.214	**-0.272**	0.431	0.479	-5.172	19
2015	-1.240	-0.622	**-0.630**	-0.151	0.282	-13.863	22
2016	-1.240	-0.493	**-0.422**	0.948	0.509	-11.822	28
2017	-0.913	0.428	**0.454**	2.219	0.993	6.350	14
2018	-1.043	0.477	**0.552**	2.789	1.199	8.282	15
2019	-1.025	0.013	**0.140**	1.695	0.892	1.960	14
2020	-0.751	1.362	**1.074**	2.628	1.211	13.957	13
2021	-1.379	-1.201	**-0.365**	1.555	1.193	-4.383	12
GPP_{EC} (z-scores)							
Year	Min	Median	Mean*	Max	S.D	Sum	n
2014	-	-	-	-	-	-	0
2015	-	-	-	-	-	-	0
2016	-1.134	-0.442	**-0.414**	0.585	0.632	-4.973	12
2017	-1.079	-0.230	**0.073**	1.857	1.051	1.016	14
2018	-0.939	0.093	**0.354**	1.872	0.959	5.303	15
2019	-1.201	-0.499	**-0.293**	0.986	0.711	-4.104	14
2020	-1.157	0.865	**0.573**	2.625	1.174	7.445	13
2021	-1.133	-0.622	**-0.074**	2.122	1.169	-0.884	12
$MODIS_{GPP}$ (z-scores)							
Year	Min	Median	Mean*	Max	S.D	Sum	n
2014	-1.340	-0.005	**-0.123**	1.088	0.743	-2.346	19
2015	-1.433	-0.644	**-0.634**	-0.072	0.356	-13.946	22
2016	-1.433	-0.505	**0.421**	0.732	0.568	-11.781	28
2017	-1.294	0.388	**0.171**	1.598	0.967	2.391	14
2018	-1.232	0.345	**0.419**	2.758	1.323	6.278	15
2019	-1.325	-0.150	**-0.036**	1.103	0.858	-0.501	14
2020	-0.907	1.134	**0.885**	2.727	1.190	11.511	13
2021	-1.309	0.253	**0.361**	2.464	1.014	4.329	12

* z-scores calculated from all Greenup – Dormancy dekads in all years, therefore the mean is not 0.

* A mean value below 0 indicates a full season where growth could be considered weak compared to the long-term average.

Phenology summarised in Table 2.

4.2 Correlation Analysis

To test the strength and statistical significance of the various pairs of datasets, the Pearson's correlation test (under assumption of Gaussian distribution) was conducted with the results summarised in Table 6, and the linear relationships being graphically depicted in Appendix A.

Table 6 – Results of Correlation Analysis of the Full and Anomaly Analysis Datasets

Variables	t	df	p-value	Pearson's Correlation Co-efficient (r)	Confidence Interval (95%)		n
Full Datasets (Including All Phenological Phases)							
*f*APAR to GDMP	14.804	183	<2.2e-16	0.738	0.665	0.798	185
$MODIS_{GPP}$ to *f*APAR	6.850	187	1.035e-10	0.448	0.326	0.555	189
$MODIS_{GPP}$ to GDMP	20.021	282	<2.2e-16	0.766	0.713	0.810	284
*f*APAR to GPP_{EC}	9.050	187	<2.2e-16	0.552	0.444	0.644	189
GDMP to GPP_{EC}	18.466	211	<2.2e-16	0.786	0.729	0.832	213
$MODIS_{GPP}$ to GPP_{EC}	26.862	218	<2.2e-16	0.876	0.842	0.904	220
Anomaly (z-scores) Analysis Dataset							
(z)*f*APAR to (z)GDMP	5.250	69	1.6e-06	0.534	0.344	0.682	71
(z)$MODIS_{GPP}$ to (z)*f*APAR	0.907	69	0.368	0.109	-0.128	0.333	71
(z)$MODIS_{GPP}$ to (z)GDMP	10.837	125	<2.2e-16	0.696	0.594	0.776	127
(z)*f*APAR to (z)GPP_{EC}	2.047	69	0.045	0.239	0.006	0.448	71
(z)GDMP to (z)GPP_{EC}	8.108	82	4.3e-12	0.667	0.528	0.771	84
(z)$MODIS_{GPP}$ to (z)GPP_{EC}	14.575	82	<2.2e-16	0.849	0.776	0.899	84

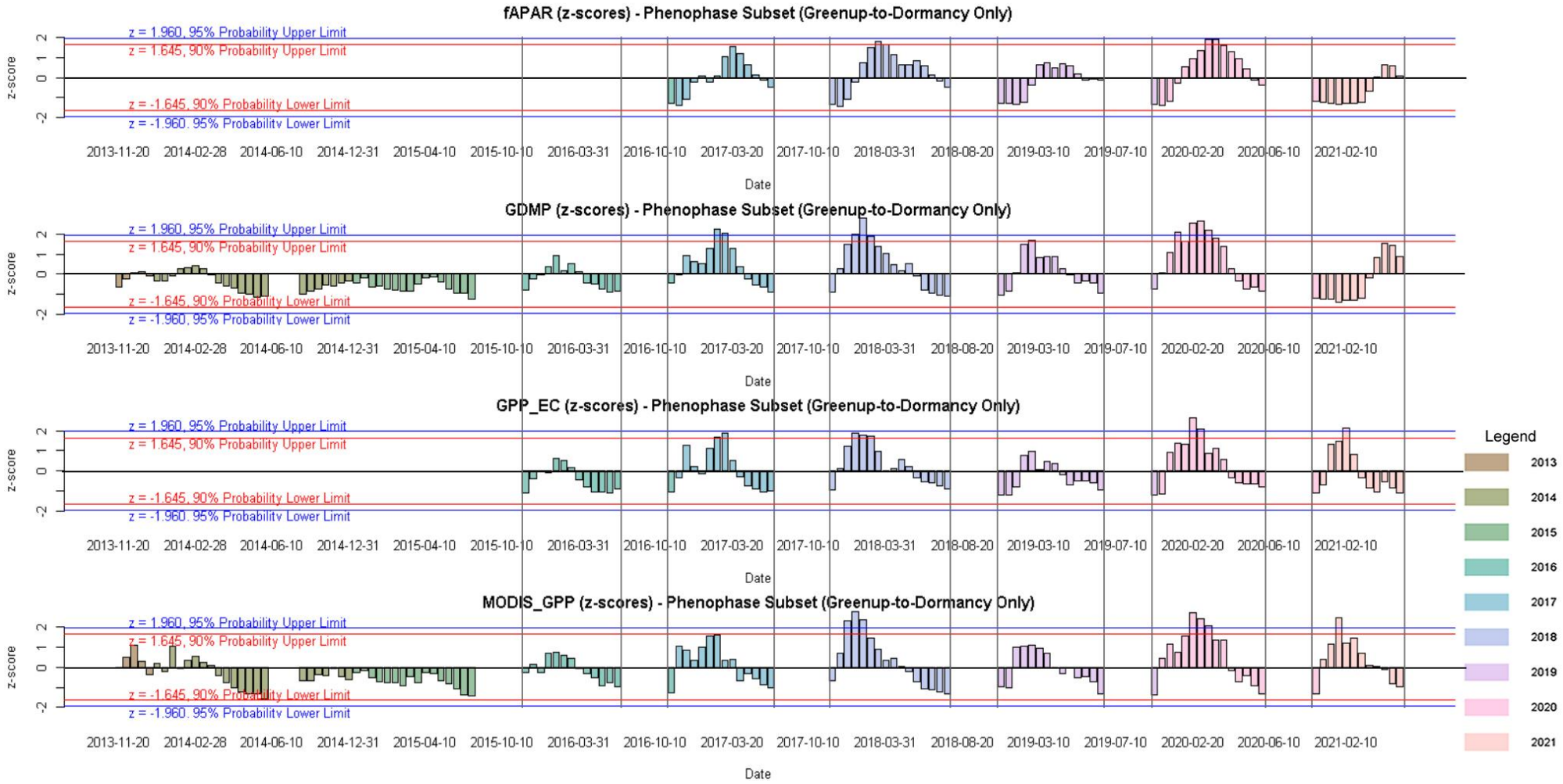


Figure 11: Bar Graphs of for Individual Growing Seasons (z-scores of fAPAR, GDMP, GPP_{EC} and $MODIS_{GPP}$ (Phenophase Subset *Datasets) - (Note: $MODIS_{GPP}$ are 8-day averages, all others 10-day averages.)*

Boxplots of Dataset Values by Month

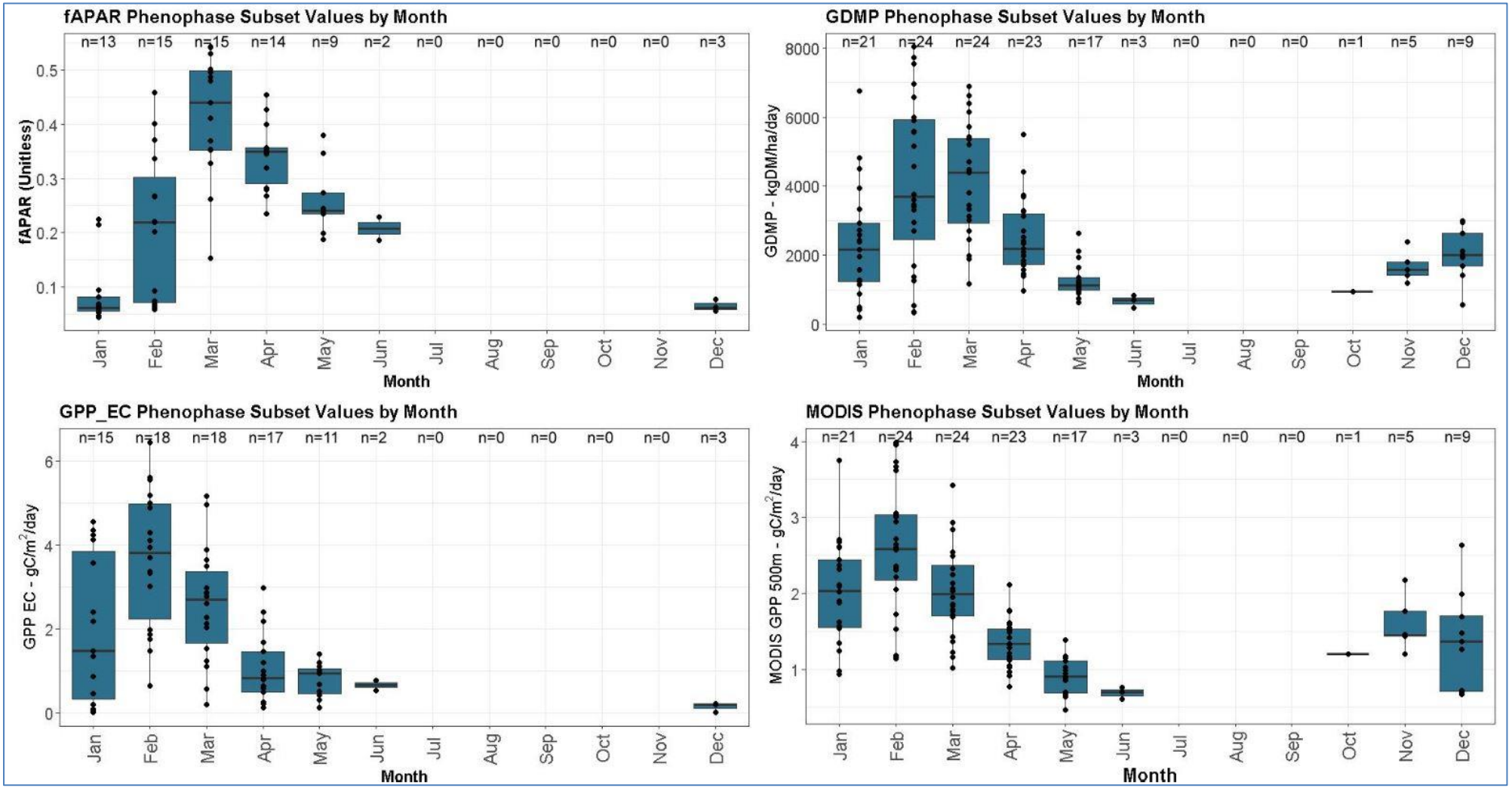


Figure 12 – Boxplots of Phenophase Datasets by Month

Boxplots of Dataset z-scores by Month

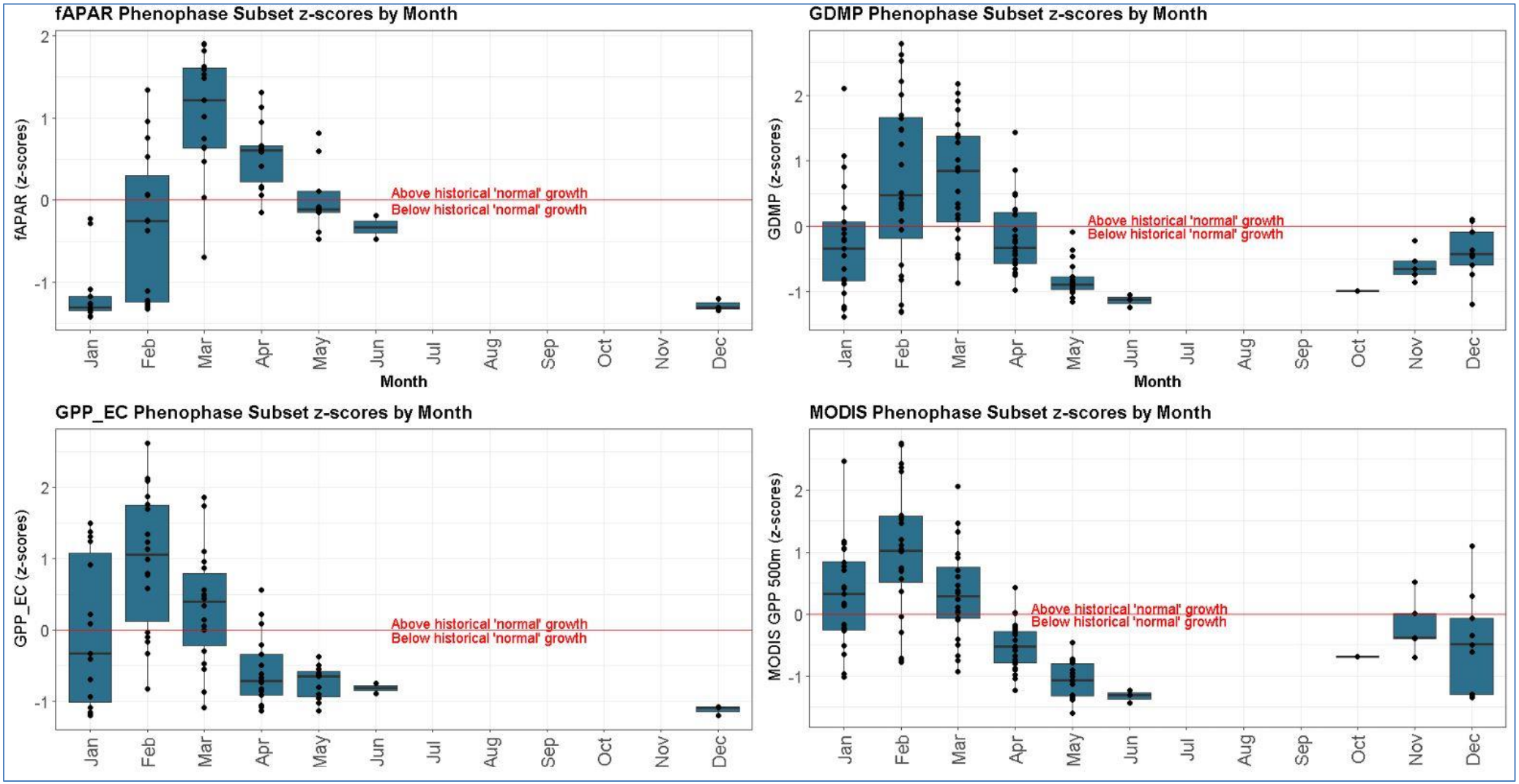


Figure 13 – Boxplots of z-scores of Phenophase Datasets *by Month*

Boxplots of Dataset Values by Year

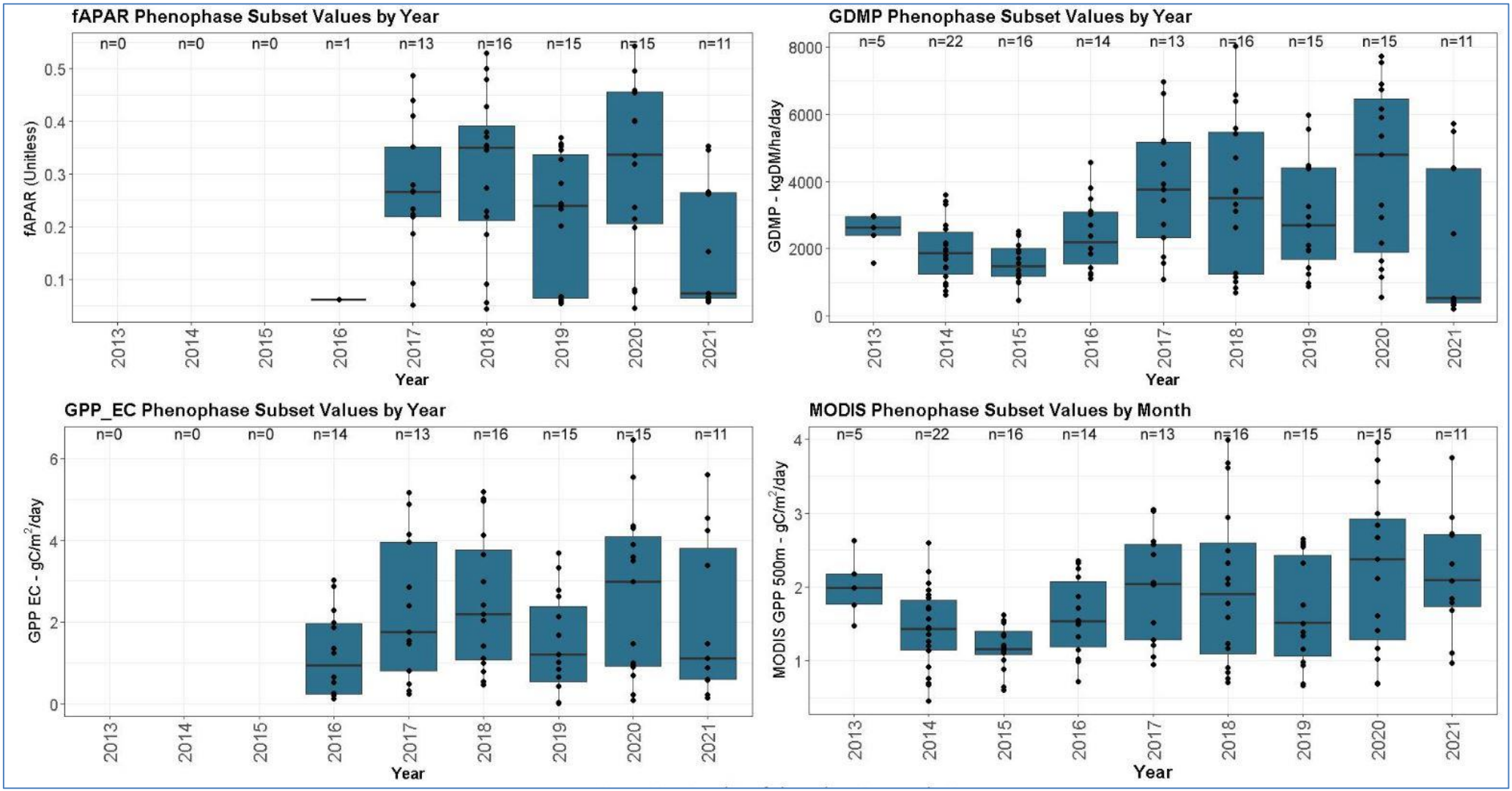


Figure 14 – Boxplots of Phenophase Datasets by Year

Boxplots of Dataset z-scores by Year

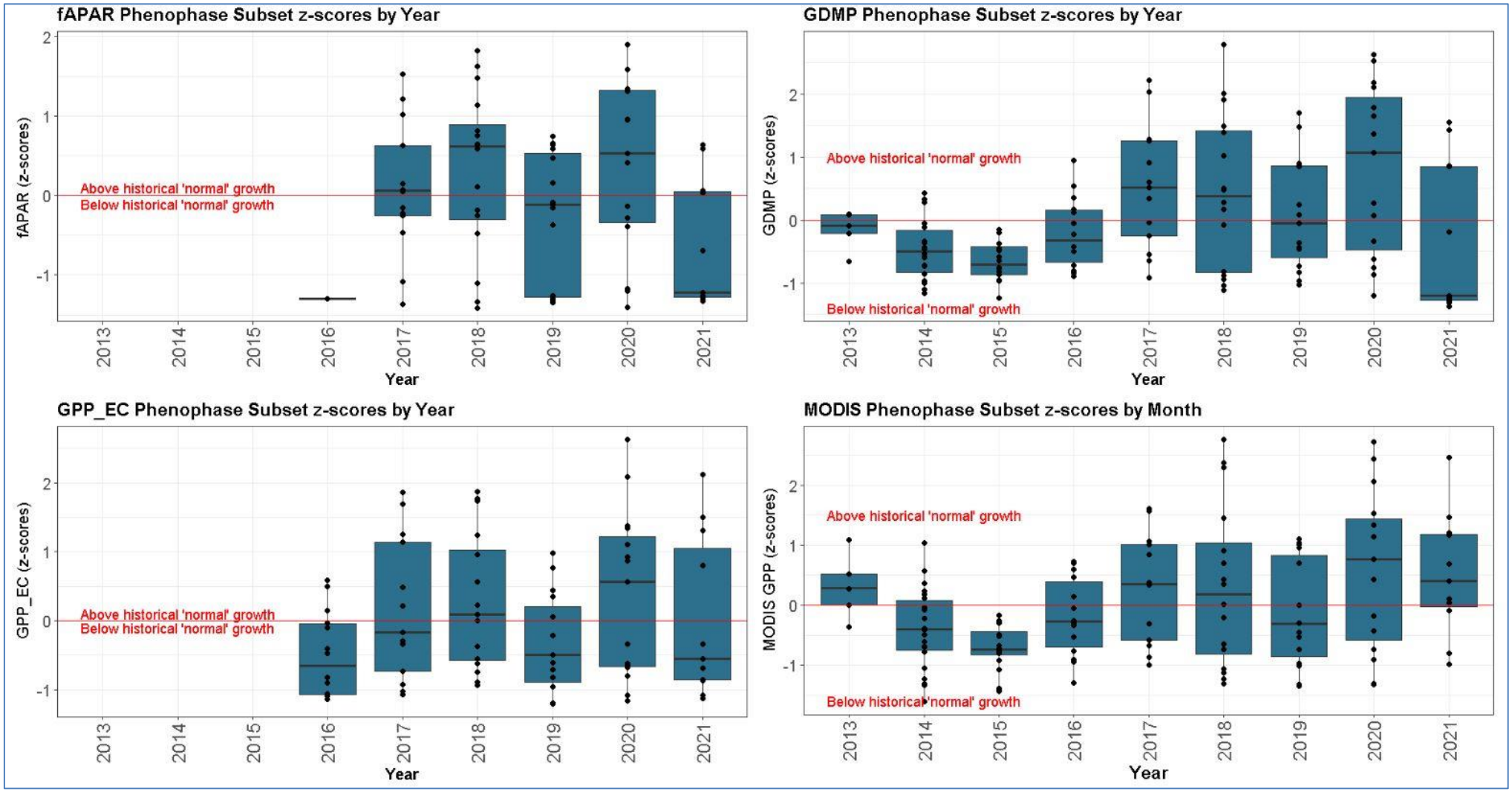


Figure 15 – Boxplots of z-scores of Phenophase Datasets by Year

4.3 Linear Regression

Despite weak correlations being discovered during the correlation analysis, it was decided to also conduct linear regression (under assumption of Gaussian distribution) on the various anomaly data pairs. This decision was made for a few reasons, namely to apply a precautionary approach by double-checking and verifying the collation of the datasets, and the findings of the correlation analysis, and to be able to derive metrics of the Correlation Coefficient (r), and the Coefficient of Determination (R^2), since other literature that was reviewed, used linear regression, and reported in the R^2 metric. Additionally, linear regression delivers equations describing the relationship between dependent and independent variables, which are sometimes useful (Box, 1976, 1979) for making predictions. For this further assessment of *f*APAR, GDMP and $MODIS_{GPP}$, as indicators of GPP_{EC} (the dependent variable), Ordinary Least Squares (OLS) linear regression was undertaken using the "jtools" package (Long, 2022) and the "lm" function in R (R Core Team, 2021). Linear regression can be represented by equation 5, and seeks to define a linear equation that describes the observed relationships between dependent and independent variables with the least degree of error across the entire dataset.

$$y = b0 + b1x + e \quad \textbf{Eq. 5.}$$

where:

y is the dependant variable ($GPP_{(EC)}$ – $gC/m^2/day$),

x is the remote sensed *f*APAR, GDMP, or $MODIS_{GPP}$ value,

b0 is the intercept of y, when x = 0,

b1 is the slope of the regression line,

e is the error term, or residual error.

4.3.1 Test for Linear Correlation - The first stage in the linear regression process was conducted during the Correlation Test, where it was determined that linear relationships did not exist for all biophysical indicators (Table 6 and Appendix A). The strengths of the correlations with GPP_{EC} (Table 6) were weak for *f*APAR (r = 0.239, p=0.045), strong for GDMP (r = 0.667, p = 4.3e-12) and strongest for $MODIS_{GPP}$ (r = 0.849, p<2.2e-16), however all these variable pairs were subjected to linear regression analysis under assumption of Gaussian distribution.

4.3.2 Test for Normality – The datasets were tested for normality using the Shapiro-Wilk test, where results from this test of below 0.05 would indicate that the data was not normally distributed. Shapiro-Wilk results (w-value) (Table 7) for the phenophase subset dataset for *f*APAR, GDMP, $MODIS_{GPP}$ and GPP_{EC} were W=0.933, p<0.001; W=0.915, p<0.0001, W=0.911, p<0.0001 and W=0.947, p<0.0001, respectively.

4.3.3 Test for Outliers – The next stage in the linear regression process is usually to check the data for outliers, and remove any datapoints that might be erroneous artefacts from the sensor, data collection process, or other error. For this study however, the purpose is to detect anomalies within the dataset and removing outliers, unless being obvious erroneous artefacts, would effectively remove the data that was being sought for testing. Nevertheless, the data was tested for any obvious erroneous observations, and none were observed, therefore all data points were accepted as being valid, and used in the Linear Regression. Boxplots of the data (Appendix B) were generated to visually assess for any erroneous artefacts in each dataset.

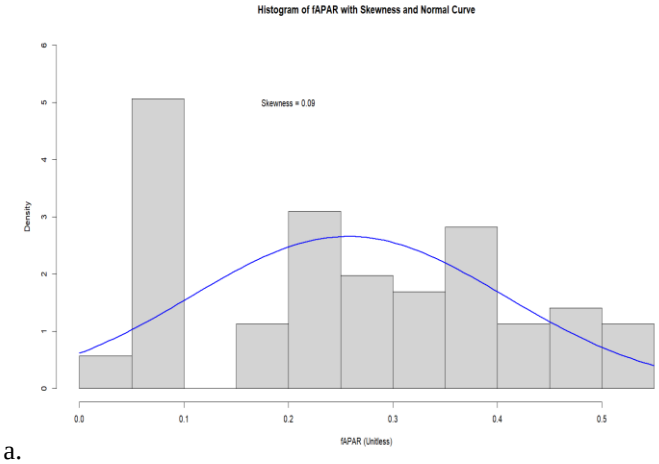
It is noteworthy however, that vegetation growth early warning systems typically define the threshold at which an observation is determined to be an anomaly, such as defined by Rembold *et al.* (2017) who used a threshold of -1 (standard deviation) to define criticality (Rembold *et al.*, 2017). It is also noteworthy, that although positive anomalies (> 0 S.D) are useful for indication of above normal vegetation growth, they are not typically used in early warning systems since these systems are designed to detect critical negative growth, as threats to food security.

4.3.4 Tests for Skewness

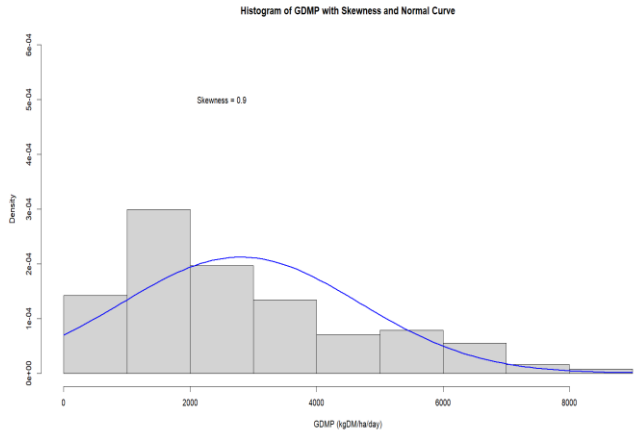
Although Gaussian distribution was assumed, the various datasets were tested for skewness, or how far each dataset was skewed away from a Gaussian distribution. A skewness value of zero indicates that the tails of the data are equally balanced on either side of the mean of the dataset. In this study the range of skewness for assumption of normal distribution was as described by Koh (2014), being a skewness value range between -2 and +2. The results (Table 7, Figures 16 a, b, c and d) indicate that *f*APAR (skewness = 0.09) displayed the closest distribution to a normal distribution, followed by GPP_{EC} (skewness = 0.70), then $MODIS_{GPP}$ (skewness = 0.74), lastly GDMP (skewness = 0.90).

Table 7 – Results of Normal Distribution and Skewness Tests

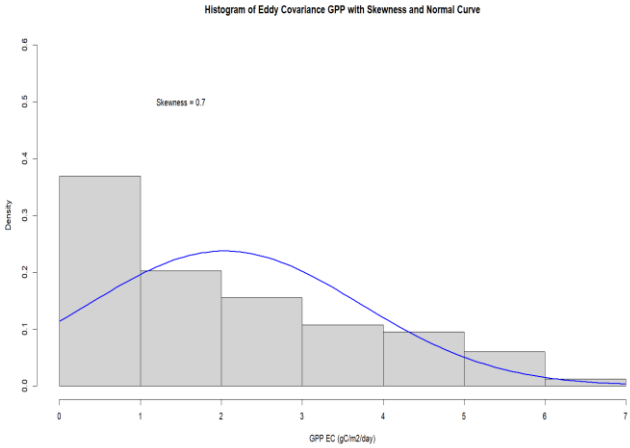
Analysis Datasets	w-values	Normality Test Result	Skewness Result
*f*APAR	0.933	Pass	0.09
GDMP	0.915	Pass	0.90
GPP_{EC}	0.911	Pass	0.70
$MODIS_{GPP}$	0.947	Pass	0.74



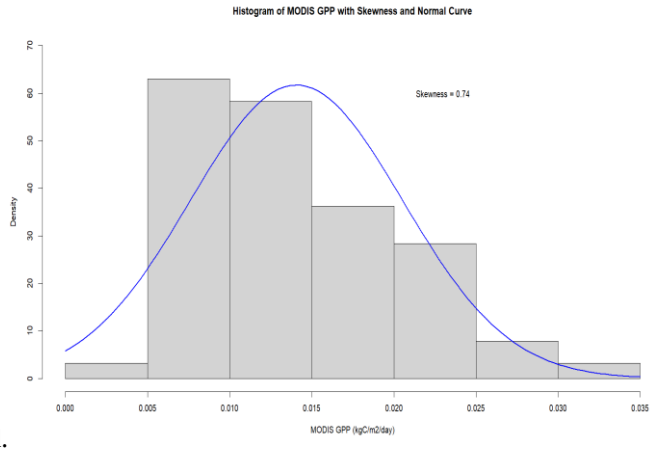
a.



b.



c.



d.

Figure 16 – Histograms of: a. fAPAR, b. GDMP, c. GPP_{EC}, d. $MODIS_{GPP}$

4.3.5 GPP_{EC} ~ *f*APAR Linear Regression

Table 8 –Linear Regression – GPP_{EC} ~ fAPAR Results

Residuals - GPP ($gC/m^2/day$)					
Min	**1Qu.**	**Median**	**3Qu.**	**Max**	**n**
-1.938	-1.207	-0.642	0.799	3.313	44
Coefficients					
	Estimate	**Std. Error**	**t value**	**Pr (>|t|)**	**Sig.**
(Intercept)	1.012	0.484	2.092	0.043	0.01
*f*APAR (Unitless)	3.976	1.588	2.503	0.016	0.01
Residual standard error: 1.611 on 42 degrees of freedom					
Multiple R-squared: 0.129, Adjusted R-squared: 0.109					
F-statistic: 6.267 on 1 and 42 DF, p-value: 0.016					
RMSE: 1.88 (GPP - $gC/m^2/day$)					

$$GPP\ (gC/m^2/day) = 1.012 + (3.976\ x\ fAPAR_{unitless}) + 1.611 \quad \textbf{Eq. 6.}$$

4.3.5.a Predictions (GPP_{EC} ~ *f*APAR Linear Regression Equation)

The linear equation for GPP_{EC} ~ *f*APAR (Equation 6) was used to predict GPP ($gC/m^2/day$) (Figure 17.a) from *f*APAR (Unitless) remote sensed data, using a random test data subset (40%) of the dataset. The greyed area indicates the 95% confidence interval and displays a relatively larger uncertainty of the regression, when compared to GDMP and $MODIS_{GPP}$.

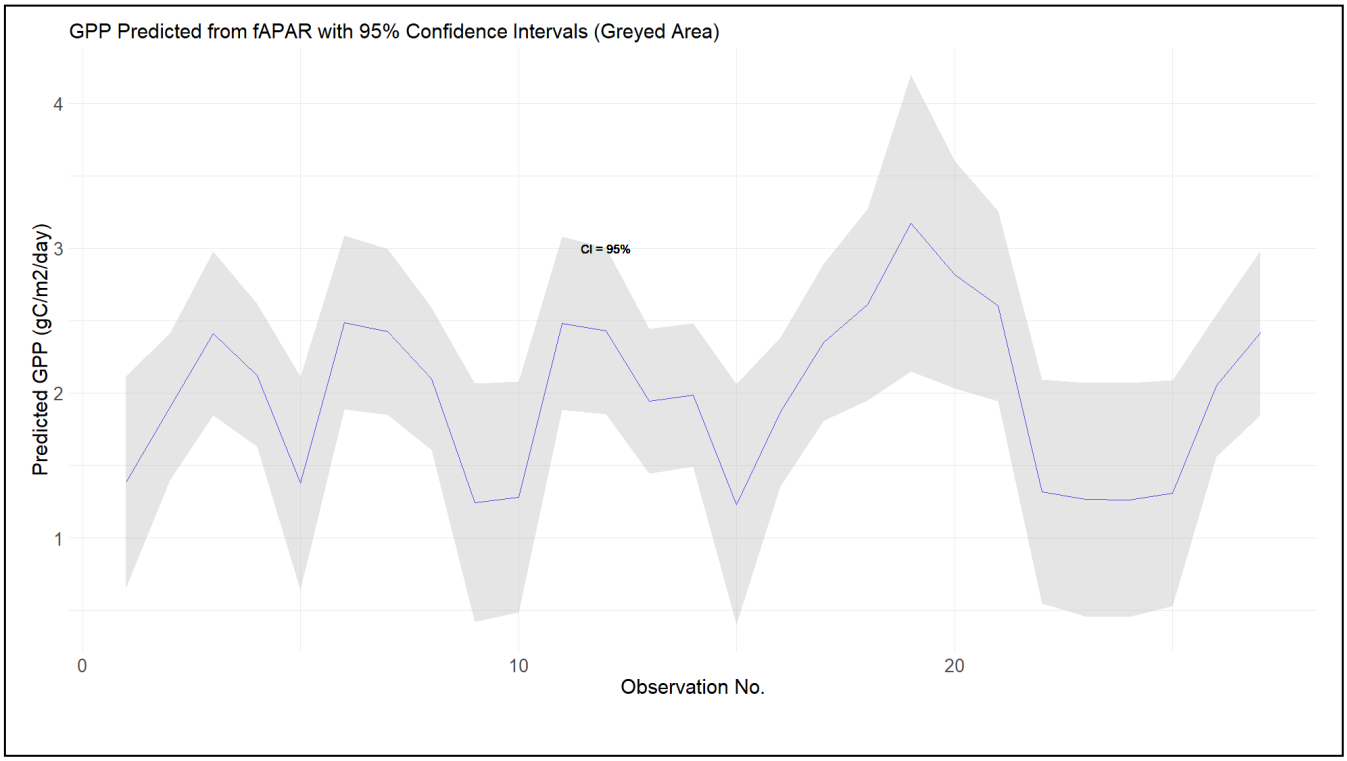


Figure 17.a – Linear Regression Predictions of GPP ($gC/m^2/day$) from fAPAR

4.3.5.b Residuals of the *f*APAR ~ GPP_{EC} Linear Regression Equation

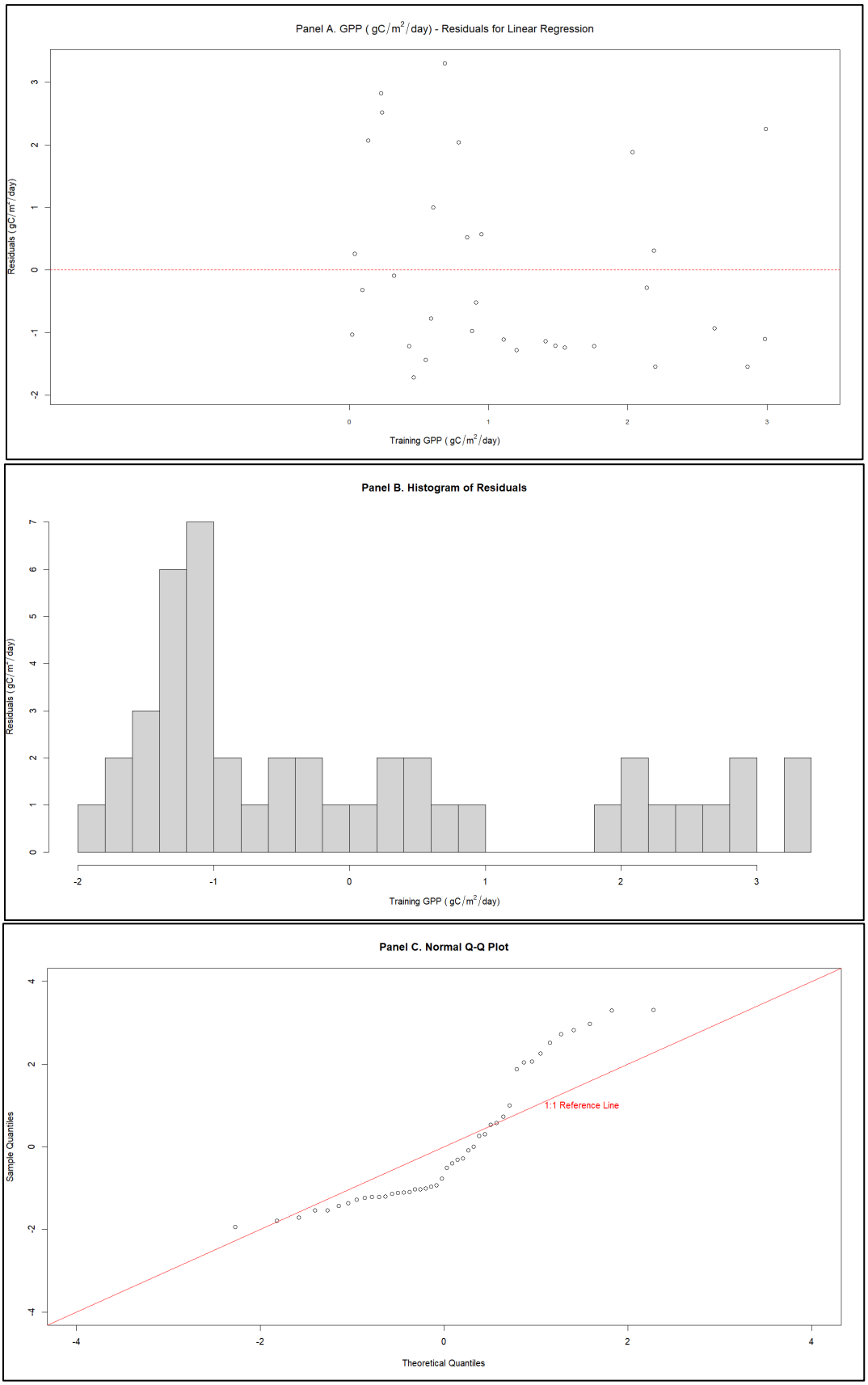


Figure 17.b – Graphic Representations of Residuals of the GPP ~ fAPAR Linear Regression Equation

In Panel A (Fig. 17.b) the x-axis indicates the GPP values ($gC/m^2/day$) in the test dataset, and the y-axis indicates the residuals, or how far, in $gC/m^2/day$, the predicted values lie from these values. The closer the residuals are to the horizontal red line at zero (Panel A), the more accurately the model performs. The residuals displayed in the scatterplot, indicate that the GPP_{EC} ~ *f*APAR regression displays a tendency to predict GPP ($gC/m^2/day$) with a relatively higher degree of inaccuracy (RMSE = 1.88 $gC/m^2/day$) than GDMP (RMSE = 1.34 $gC/m^2/day$) and $MODIS_{GPP}$ (RMSE = 0.599 $gC/m^2/day$), since the residuals are not close to the horizontal red line, and the RMSE is the highest of the three regressions assessed.

Additionally, the histogram (Panel B) of the residuals indicates the distribution of the residuals of the prediction model. The histogram can be used to assess the residuals for normality, or the skewed nature of the distributions, and assists in determining if the model is normally distributed, or biased high or low. A histogram centred around the zero, and having an equal number of residuals on either side of the zero, would indicate a model that was normally distributed, with no positive, nor negative bias. The histogram of the residuals of the GPP_{EC} ~ *f*APAR regression indicate that the model is negatively biased (low) with the majority (n = 27) of residuals below zero, and a minority (n=17) above zero.

In the Normal Q-Q Plot (Panel C) the red sloping line is a 1:1 reference line, representing a theoretical, perfectly fit (r = 1) model. This reference line assists in visually assessing the model's performance. The tendency for the GPP_{EC} ~ *f*APAR regression to predict lower than actual is evident in the residuals plotted below the reference line, whilst the model still displays a potential to over predict, on the higher end (> 1) of the scale.

4.3.6 GPP_{EC} ~ GDMP Linear Regression

Table 9 – Linear Regression –GPP_{EC} ~ GDMP Results

Residuals GPP (gC/m²/day)					
Min	**1Qu.**	**Median**	**3Qu.**	**Max**	**n**
-2.482	-0.651	-0.096	0.421	5.205	52
Coefficients					
	Estimate	**Std. Error**	**t value**	**Pr (>|t|)**	**Sig.**
(Intercept)	1.341e-01	3.134e-01	0.428	0.671	>0.05
GDMP (kgDM/ha/day)	6.022e-04	7.901e-05	7.623	6.4e-10	<0.001
Residual standard error: 1.223 on 50 degrees of freedom					
Multiple R-squared: 0.538, Adjusted R-squared: 0.528					
F-statistic: 58.11 on 1 and 50 DF, p-value: 6.399e-10					
RMSE: 1.34 (GPP - gC/m²/day)					

$$GPP\,(gC/m^2/day) = 1.341e\text{-}01 + (6.022e\text{-}04 \times GDMP\,(kgDM/ha/day)) + 1.223 \quad \textbf{Eq. 7.}$$

4.3.6.a Predictions (GPP_{EC} ~ GDMP Linear Regression Equation)

The linear equation for GPP_{EC} ~ GDMP (Equation 7) was used to predict GPP (gC/m²/day) (Figure 18.a) from remote sensed GDMP (kgDM/ha/day), using a random test data subset (40%) of the dataset. The greyed area indicates the 95% confidence interval and displays a relatively lower uncertainty of the regression, than the *f*APAR-based regression.

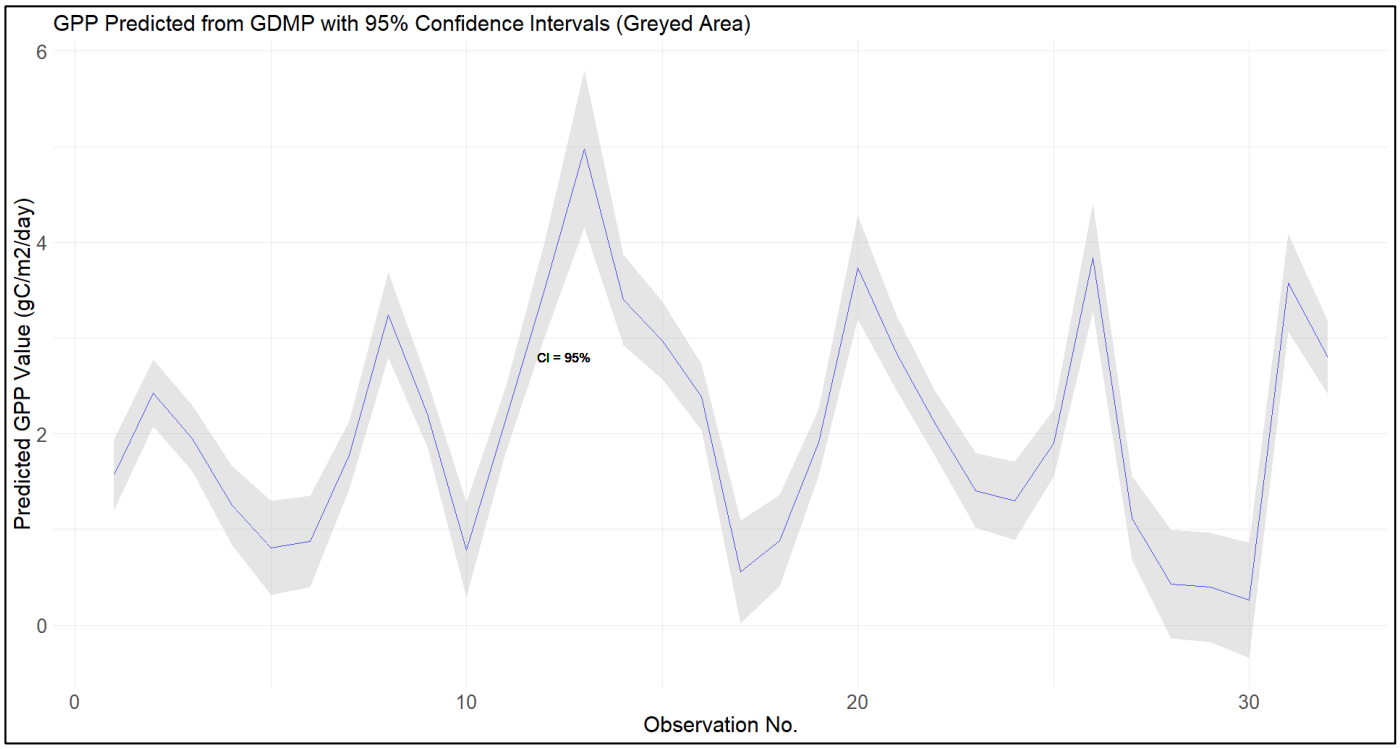


Figure 18.a – Linear Regression Predictions of GPP (gC/m²/day) from GDMP

4.3.6.b Residuals of the GPP_{EC} ~ GDMP Linear Regression Equation

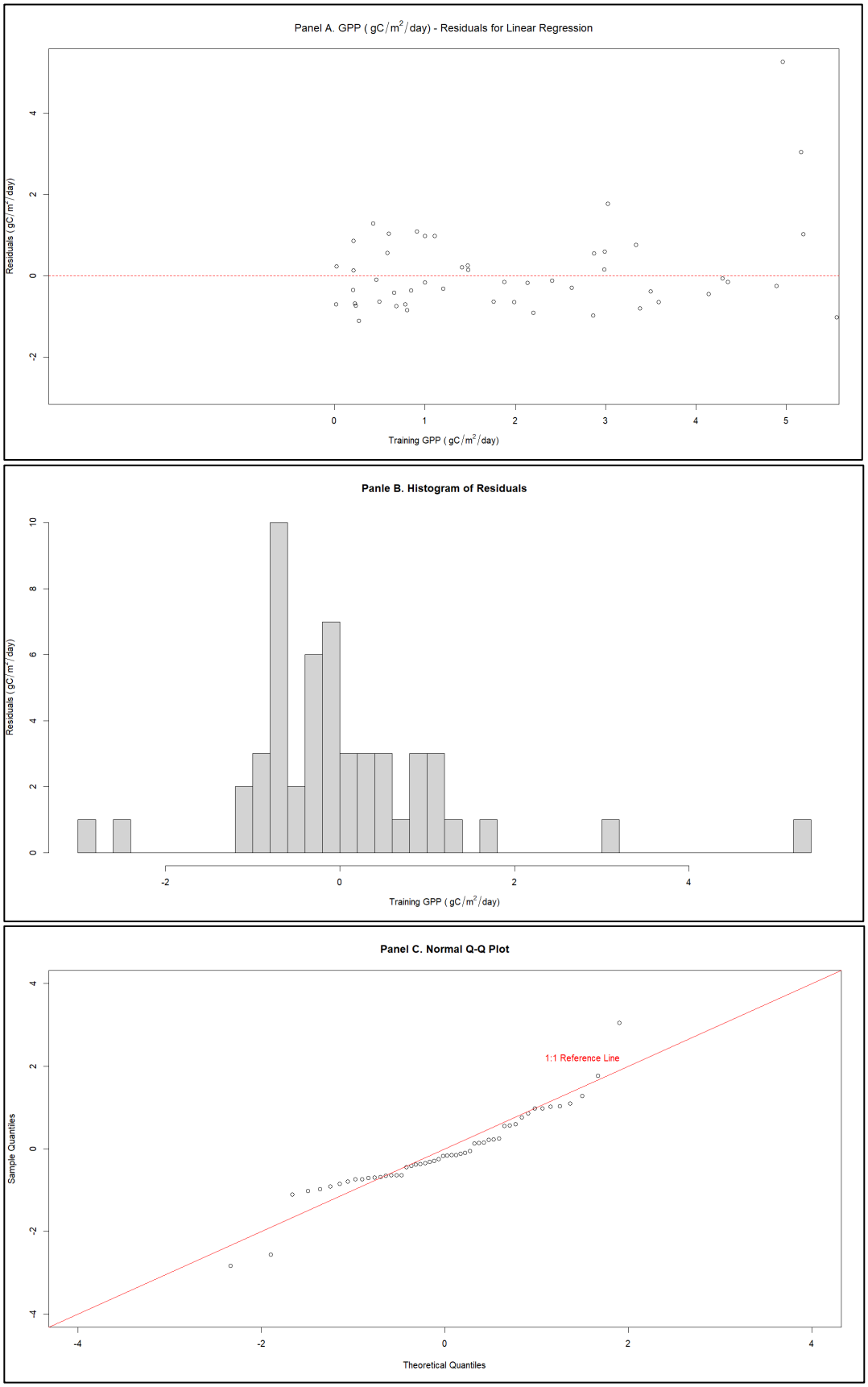


Figure 18.b – Graphic Representations of Residuals of the GPP ~ GDMP Linear Regression Equation

In Panel A (Fig. 18.b), the residuals displayed in the scatterplot, indicate that the GPP_{EC} ~ GDMP regression displays the residuals to be closer to the red horizontal line at zero than seen for *f*APAR, with a tendency to predict GPP ($gC/m^2/day$) slightly more accurately (RMSE = 1.34 $gC/m^2/day$), than the GPP_{EC} ~ *f*APAR regression (RMSE = 1.88 $gC/m^2/day$).

Additionally, the histogram (Panel B) of the residuals of the GPP_{EC} ~ GDMP regression indicate that the model is also biased low with the majority (n = 32) of residuals below zero, and a minority (n=20) above zero.

In the Normal Q-Q Plot (Panel C) the tendency for the GPP_{EC} ~ GDMP regression to predict slightly lower than actual is evident in the residuals plotted below the 1:1 reference line, whilst the model still displays a potential to over predict to a lesser degree, on the higher end (>1) of the scale.

4.3.7 GPP_{EC} ~ $MODIS_{GPP}$ Linear Regression

Table 10 – Linear Regression – GPP_{EC} ~ $MODIS_{GPP}$ Results

Residuals (gC/m²/day)					
Min	**1Qu.**	**Median**	**3Qu.**	**Max**	**n**
-2.447	-0.529	0.002	0.549	2.261	52
Coefficients					
	Estimate	**Std. Error**	**t value**	**Pr (>|t|)**	**Sig.**
(Intercept)	-1.155	0.286	-4.043	0.0001	<0.001
$MODIS_{GPP}$ (gC/m²/day)	1.652	0.136	12.173	<2e-16	<0.001
Residual standard error: 0.830 on 50 degrees of freedom					
Multiple R-squared: 0.748, Adjusted R-squared: 0.743					
F-statistic: 148.2 on 1 and 50 DF, p-value: <2.2e-16					
RMSE: 0.599 (GPP - gC/m²/day)					

$$GPP_{EC} = -1.155 + (1.652 \times MODIS_{GPP}\ (gC/m^2/day)) + 0.830 \quad \textbf{Eq. 8.}$$

4.3.7.a Predictions (GPP_{EC} ~ $MODIS_{GPP}$ Linear Regression Equation)

The linear equation for GPP_{EC} ~ $MODIS_{GPP}$ (Equation 8) was used to predict GPP (gC/m²/day) from remote sensed $MODIS_{GPP}$ (Figure 19.a), using a random test data subset (40%) of the dataset. The greyed area (95% confidence interval) displays the lowest uncertainty of the three regressions assessed.

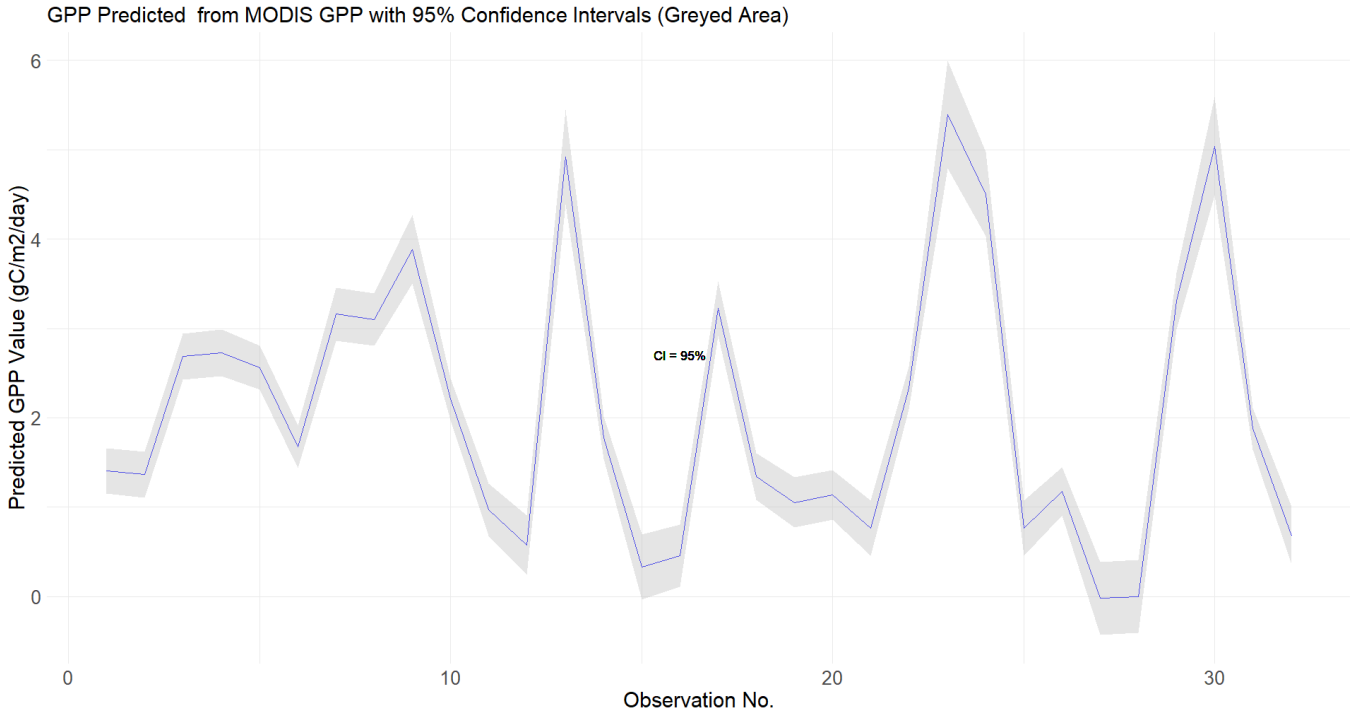


Figure 19.a – Linear Regression Predictions of GPP (gC/m²/day) from $MODIS_{GPP}$

4.3.7.b Residuals of the GPP_{EC} ~ $MODIS_{GPP}$ Linear Regression Equation

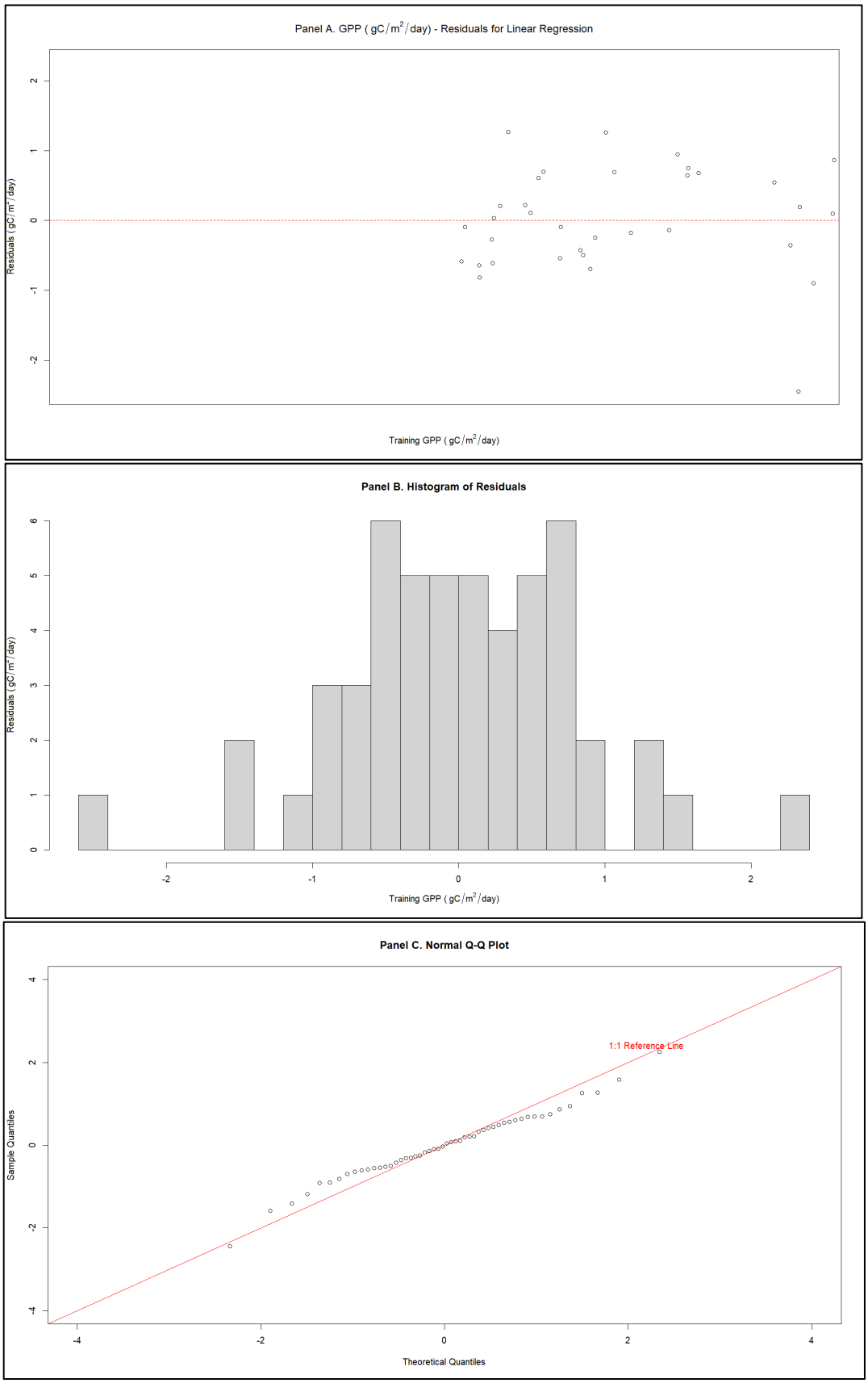


Figure 19.b – Graphic Representations of Residuals of the GPP ~ $MODIS_{GPP}$ Linear Regression Equation

In Panel A (Fig. 19.b), the residuals displayed in the scatterplot, indicate that the GPP_{EC} ~ $MODIS_{GPP}$ regression displays a tendency to predict GPP ($gC/m^2/day$) more accurately (RMSE = 0.599 $gC/m^2/day$), than the GPP_{EC} ~ *f*APAR (RMSE = 1.88 $gC/m^2/day$), and the GPP_{EC} ~ GDMP (RMSE = 1.34 $gC/m^2/day$) regressions.

Additionally, the histogram (Panel B) of the residuals of the GPP_{EC} ~ $MODIS_{GPP}$ regression indicates that the model that results in residuals that is almost a Gaussian distribution, with an even split (n = 26) of residuals below zero, and (n=26) above zero. Most of the residuals (n = 44) fall between -1.0 and 1.0, indicating that the model closely predicts GPP ($gC/m^2/day$) from $MODIS_{GPP}$ observations.

In the Normal Q-Q Plot (Panel C) the GPP_{EC} ~ GDMP regression closely follows the 1:1 reference line, again indicating the strength of the model in predicting GPP ($gC/m^2/day$) from $MODIS_{GPP}$ observations.

4.4 Hypotheses Testing

The data tested, suggests a weak, and statistically significant (r = 0.534, p = 1.6e-06, Table 6) correlation between standard scores of (z)*f*APAR and (z)GDMP. The null hypothesis relating to (z)*f*APAR and (z)GDMP is thus not rejected.

The data tested, suggests a weak, and statistically insignificant (r = 0.239, p = 0.045, Table 6) correlation between standard scores of (z)*f*APAR and (z)GPP_{EC}. The null hypothesis relating to (z)*f*APAR and (z)GPP_{EC} is thus not rejected.

The data tested, suggests a strong and statistically significant (r = 0.667, p = 4.3e-12, Table 6) correlation between standard scores of (z)GDMP and (z)GPP_{EC}. The null hypothesis relating to (z)GDMP and (z)GPP_{EC} is thus rejected.

The data tested, suggests a weak, and statistically insignificant (r = 0.109, p = 0.365, Table 6) correlation between standard scores of (z)$MODIS_{GPP}$ and (z)*f*APAR. The null hypothesis relating to (z)$MODIS_{GPP}$ and (z)*f*APAR is thus not rejected.

The data tested, suggests a strong, and statistically significant (r = 0.696, p = <2.2e-16, Table 6) correlation between standard scores of (z)$MODIS_{GPP}$ and (z)GDMP. The null hypothesis relating to (z)$MODIS_{GPP}$ and (z)GDMP is thus rejected.

The data tested, suggests a strong, and statistically significant (r = 0.849, p = <2.2e-16, Table 6) correlation between standard scores of (z)$MODIS_{GPP}$ and (z)GPP_{EC}. The null hypothesis relating to (z)$MODIS_{GPP}$ and (z)GPP_{EC} is thus rejected.

5. Discussion

The remote sensed biophysical indicators, *f*APAR, GDMP and $MODIS_{GPP}$, use different sensors; input data; and algorithms, to represent the state of, and fluxes observed in, vegetation growth (Huete, 1999; Martins, Trigo and Freitas, 2020c, 2020a). Although these indicators differ in their approaches, they all attempt to represent vegetation growth dynamics, and hence it stands to reason that they would each have close correlations to ground-based vegetation growth dynamics observations of another vegetation growth indicator, namely GPP, such as derived from EC measurements. With spatially limited EC monitoring stations, these remote sensed biophysical indicators can provide an efficient system of monitoring GPP for anomalous growth, using z-scores, which can be used for crop management and food security. However, the performance of any such detection system, is affected by the performance of the chosen remote sensed biophysical indicator, regarding how well the remote sensed indicator temporally mimics actual GPP, and therefore the choice of the remote sensed indicator is of importance.

In my semi-arid grassland study, the performance of three remote sensed indicators has been tested, with my findings indicating that $MODIS_{GPP}$ ($r = 0.849$, $p < 2.2e\text{-}16$, Table 6 and $R^2 = 0.748$, RMSE=0.599 $gC/m^2/day$, Table 10) performed better than GDMP ($r = 0.667$, $p = 4.3e\text{-}12$, Table 6 and $R^2 = 0.538$, RMSE=1.34 $gC/m^2/day$, Table 9), which performed better than *f*APAR ($r = 0.239$, $p = 0.045$, Table 6 and $R^2 = 0.129$, RMSE = 1.88 $gC/m^2/day$, Table 8). These findings agree with Zhang *et al.* (2021), who found that $MODIS_{GPP}$ strongly agreed ($r^2 = 0.73$, RMSE = 1.14 $gC/m^2/day$) with GPP_{EC}, and that there was also strong agreement ($r^2 = 0.74$, RMSE = 1.90 $gC/m^2/day$) between a *GDMP type*, Sentinel-3 *f*APAR combined with meteorology data and light use efficiency (LUE) models (Zhang *et al.*, 2021), similar to how GDMP is derived from *f*APAR (Martins, Trigo and Freitas, 2020a).

It is noteworthy, however, that these strong associations were found for a northern-latitude woody savanna biome at an AmeriFlux site, Tonzi Ranch (Zhang *et al.*, 2021), whilst in a grassland biome located at the AmeriFlux site at Vaira Ranch-Ione (Zhang *et al.*, 2021), $MODIS_{GPP}$ ($r^2 = 0.49$, RMSE = 1.60 $gC/m^2/day$) performed less well than a *GDMP type*, Sentinel-3 *f*APAR based $GPP_{EC\text{-}LUE}$ ($r^2 = 0.67$, RMSE = 2.50 $gC/m^2/day$), which is in disagreement with my findings for $MODIS_{GPP}$ and GDMP in grasslands. The findings of weaker GDMP *type*, *f*APAR based associations with GPP_{EC} in grasslands, compared to woody savanna, are supported in another Zhang *et al.* (2020) study, which found greater agreement of GDMP *type*, Sentinel-3 *f*APAR based GPP with GPP_{EC}, at the Tonzi Ranch woody savanna ($r^2 = 0.65$, RMSE = 1.39 $gC/m^2/day$), than at the grassland, Kansas Field Station ($r^2 = 0.45$, RMSE = 5.54 $gC/m^2/day$) (Zhang, Zhao and Lin, 2020).

My findings, for the semi-arid grassland site, suggest that *f*APAR on its own is not sufficient to represent GPP, and considering that GDMP is derived from *f*APAR and other inputs, it is thought that the inclusion of these other inputs, being ECMWF modelled meteorological data, biome specific land cover information, and the UN LCCS's data, are responsible for the better performance of GDMP, and that these, and possibly other variables are required to fully describe GPP. Additionally, since GPP_{EC} includes vapour pressure deficits (VPD), and temperature sensitivity of respiration (Lasslop *et al.*, 2010), and *f*APAR represents the standing state of vegetation, there is likely to be a *f*APAR response lag, which is likely contributing to the relatively poor performance of *f*APAR i.e., existing green vegetation, with high *f*APAR values, would respond to diminishing VPD and changes in temperature later than indicated by GPP_{EC}.

This hypothesis, that more variables other than *f*APAR alone are required to describe GPP is confirmed in several studies. For example, a study by Stocker *et al.* (2020), described GPP as the product of APAR and LUE (GPP = PAR x *f*APAR x LUE) , and was used to define an optimality-based GPP model, the "P-model", which incorporated variables such as temperature, water vapour deficit, atmospheric pressure, and ambient CO_2 concentration, to improve simulated LUE, and hence improve (R^2 = 0.75, RMSE = 1.96 $gC/m^2/day$) predictions of GPP (Stocker *et al.*, 2020). Another example, in their recent study, Zhang *et al.* (2022) found that the P-model outperformed LUE models when predicting GPP (R^2 = 0.75, RMSE = 1.77 $gC/m^2/day$) and assessed the effectiveness of different, energy partitioning and evapotranspiration (ET) based water stress factors, into the P-model, finding the former performed better than ET-based ones (Zhang *et al.*, 2022). Additionally, a study by Stocker *et al.* (2018) concluded that fractional reduction due to soil moisture (*f*LUE), as a separate variable from vapour pressure deficit (VPD), captured substantial drought impacts better than VPD alone, and further concluded that "*f*LUE reductions are largest in drought-deciduous vegetation, including grasslands" (Stocker *et al.*, 2018). From these studies it is further evidence that *f*APAR requires additional variables to describe GPP with statistical accuracy.

It has been reported that regression-based models are site, sensor; or season specific, and that the transferability of these models between sites is thus limited (Ardö *et al.*, 2022). The sensor variability appears to hold true for the current study, based on the results and variable linear regression performances observed. However, with this study being conducted at a single site, and within growing seasons only, both of which are limitations of the study, the site and seasonal transferability of the regressions obtained in this study, has not been assessed. This is not to say that the regression-based algorithms presented in this study are of no use at all. With due caution and understanding of their limitations and uncertainties, the regressions appearing in this study, particularly $MODIS_{GPP}$, could be used as a fast and simple predictor of Nama-Karoo type semi-arid grassland biomass availability, for grazing management.

6. Conclusion

The findings of the study suggest high variability between the remote sensed indicators of *f*APAR, GDMP and $MODIS_{GPP}$ in representing GPP_{EC}, at a semi-arid grassland site in South Africa. This can be seen from the results where the strongest correlation was seen between the reference data (GPP_{EC}) and $MODIS_{GPP}$ (r=0.849, p<2.2e-16, Table 6), followed by GDMP (0.667, p=4.3e-12, Table 6), and lastly *f*APAR (r=0.239, p=0.045, Table 6). These findings are again supported by the results of linear regression, which revealed that $MODIS_{GPP}$ (R^2 = 0.748, RMSE = 0.599 $gC/m^2/day$, Table 10) appears to represent changes in GPP_{EC} more accurately than GDMP (R^2 = 0.538, RMSE = 1.34 $gC/m^2/day$, Table 9), and lastly *f*APAR (R^2 = 0.129, RMSE = 1.88, Table 8).

The differences between z-scores for *f*APAR, GDMP, $MODIS_{GPP}$ and GPP_{EC}, is also evident within individual growing seasons (Figure 11). For example, at the start of each season, *f*APAR appeared to yield more observations of negative z-scores, than the other remote sensed indicators, and the reference data (GPP_{EC}), and appeared to have a short (2-3 dekad) lag in its response to both positive and negative changes in GPP_{EC}, compared to the other biomass indicators. This temporal response difference could be a significant flaw if *f*APAR were considered for use in an early warning anomaly detection system, because crop management responses to anomaly detection may be delayed by 2-3 (or more) dekads, compared to systems using GDMP or $MODIS_{GPP}$.

Additionally, with this finding that *f*APAR on its own, does not sufficiently represent GPP_{EC}, and that additional variables and LUE models, such as the optimised P-model, are likely to increase performance of *f*APAR based GPP predictions, anomalies within z-scores, calculated using biophysical indicators that incorporate additional variables, and not simply *f*APAR alone, such as the ESA's Sentinel-3 GDMP and NASA's $MODIS_{GPP}$, are likely to prove more useful and accurate for vegetation and crop anomaly early warning systems, in semi-arid grasslands.

It is thus recommended, that *f*APAR alone is not used, or is supplemented by additional LUE variables if it is to be used, in any early warning anomaly detection system for South African semi-arid grasslands.

It is additionally recommended that further similar studies, that include additional LUE variables, and perhaps the optimised P-model are undertaken to expand the knowledge on accuracy of vegetation and crop anomaly early warning systems in semi-arid grasslands, particularly in other similar mid-latitude ecozone's e.g., Australia, South America, and Mexico.

7. References

Ahlström, A. *et al.* (2015) 'The dominant role of semi-arid ecosystems in the trend and variability of the land CO2 sink', *Science*, 348(6237), pp. 895–899. doi: 10.1126/science.aaa1668.

Ardö, J. *et al.* (2022) 'State-of-the-art Review Report (D3)", Rangeland monitoring for Africa using Earth Observation (RAMONA), pp. 15, 31.

Bacour, C. *et al.* (2006) 'Neural network estimation of LAI, fAPAR, fCover and LAI×Cab, from top of canopy MERIS reflectance data: Principles and validation', *Remote Sensing of Environment*, 105(4), pp. 313–325. doi: 10.1016/j.rse.2006.07.014.

Baldocchi, D. D. (2003) 'Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: Past, present and future', *Global Change Biology*, 9(4), pp. 479–492. doi: 10.1046/j.1365-2486.2003.00629.x.

Baldocchi, D. D., Hincks, B. B. and Meyers, T. P. (1988) 'Measuring Biosphere-Atmosphere Exchanges of Biologically Related Gases with Micrometeorological Methods', *Ecology*, 69(5), pp. 1331–1340. Available at: http://links.jstor.org/sici?sici=0012-9658(198810)69%3A5%3C1331%3AMBEOBR%3E2.0.CO%3B2-T%5Cnpapers2://publication/uuid/73823DD4-B12F-4D9B-80B2-9199EA37E661.

Baret, F. *et al.* (2007) 'LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. Part 1: Principles of the algorithm', *Remote Sensing of Environment*, 110(3), pp. 275–286. doi: 10.1016/j.rse.2007.02.018.

Box, G. E. . (1976) 'Science and Statistics', *Source: Journal of the American Statistical Association*, 71(356), pp. 791–799.

Box, G. E. P. (1979) *Robustness in the Strategy of Scientific Model Building*, *Robustness in Statistics*. ACADEMIC PRESS, INC. doi: 10.1016/b978-0-12-438150-6.50018-2.

Durgun, Y. Ö. *et al.* (2016) 'Testing the contribution of stress factors to improve wheat and maize yield estimations derived from remotely-sensed dry matter productivity', *Remote Sensing*, 8(3), pp. 1–24. doi: 10.3390/rs8030170.

Gaberščik, A. and Murlis, J. (2011) 'The role of vegetation in the water cycle', *Ecohydrology and Hydrobiology*, 11(3–4), pp. 175–181. doi: 10.2478/v10104-011-0046-z.

GDAL/OGR Contributors (2020) '{GDAL/OGR} Geospatial Data Abstraction software Library'. Available at: https://gdal.org.

Haynes, K. D. *et al.* (2019) 'Representing Grasslands Using Dynamic Prognostic Phenology Based on Biological Growth Stages: 1. Implementation in the Simple Biosphere Model (SiB4)', *Journal of Advances in Modeling Earth Systems*, 11(12), pp. 4423–4439. doi: 10.1029/2018MS001540.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis, 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978.

https://www.ramona.earth (2022) *RAMONA - Rangeland Monitoring for Africa*. Available at: https://www.ramona.earth.

Huete, A. R. (1999) 'MODIS VEGETATION INDEX ALGORITHM THEORETICAL BASIS v3', *Environmental Sciences*, (Mod 13).

Jacobs, T. and VITO NV (2020) 'aoi_extract_gtiff'.

Kljun, N. *et al.* (2015) ‘A simple two-dimensional parameterisation for Flux Footprint Prediction (FFP)’, *Geoscientific Model Development*, 8(11), pp. 3695–3713. doi: 10.5194/gmd-8-3695-2015.

Koh, K. (2014). Univariate Normal Distribution. In: Michalos, A.C. (eds) Encyclopedia of Quality of Life and Well-Being Research. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0753-5_3109

Lasslop, G. *et al.* (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(1), pp. 187–208. doi: 10.1111/j.1365-2486.2009.02041.x.

De Lemos, H., Verstraete, M. M. and Scholes, M. (2020) ‘Parametric models to characterize the phenology of the lowveld savanna at skukuza, south africa’, *Remote Sensing*, pp. 1–39. doi: 10.3390/rs12233927.

Lieth, H. (1973) ‘Phenology in Productivity Studies’, pp. 29–46. doi: 10.1007/978-3-642-85587-0_4.

Long, J. A. (2022) ‘jtools: Analysis and Presentation of Social Scientific Data’. Available at: https://cran.r-project.org/package=jtools.

Martins, J. P., Trigo, I. and Freitas, S. C. e (2020a) ‘Copernicus Global Land Operations ”Vegetation and Energy” “CGLOPS-1” - DMP/GDMP’, *Copernicus Global Land Operations*, pp. 1–93. Available at: https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS1_ATBD_DMP300m-V1.1_I1.20.pdf.

Martins, J. P., Trigo, I. and Freitas, S. C. e (2020b) ‘Copernicus Global Land Operations ”Vegetation and Energy” “CGLOPS-1” - Dry Matter Productivity (DMP)’, *Copernicus Global Land Operations*, pp. 1–93. Available at: https://land.copernicus.eu/global/products/dmp.

Martins, J. P., Trigo, I. and Freitas, S. C. e (2020c) ‘Copernicus Global Land Operations ”Vegetation and Energy” “CGLOPS-1” - fAPAR’, *Copernicus Global Land Operations*, pp. 1–93. Available at: https://land.copernicus.eu/global/products/fapar.

Meroni, M. *et al.* (2014) ‘A phenology-based method to derive biomass production anomalies for food security monitoring in the Horn of Africa’, *International Journal of Remote Sensing*. Taylor & Francis, 35(7), pp. 2472–2492. doi: 10.1080/01431161.2014.883090.

Meroni, M. *et al.* (2019) *The warning classification scheme of ASAP – Anomaly hot Spots of Agricultural Production, v4.0, JRC Technical Report*. doi: 10.2760/798528.

Monteith (1972) ‘Solar Radiation and Productivity in Tropical Ecosystems Author (s): J . L . Monteith Source : Journal of Applied Ecology , Vol . 9 , No . 3 (Dec ., 1972), pp . 747-766 Published by : British Ecological Society Stable URL : http://www.jstor.org/stable/’, *Society*, 9(3), pp. 747–766.

NASA (2023) *Moderate Resolution Imaging Spectroradiometer - About*. Available at: https://modis.gsfc.nasa.gov/about/ (Accessed: 16 May 2023).

National Phenology Network (USA-NPN) (2023) *Definition of Phenophase, Website*. Available at: https://usanpn.org/taxonomy/term/16#:~:text=An observable stage or phase,a few days or weeks. (Accessed: 8 May 2023).

Pinker, R. T. *et al.* (2010) ‘Impact of satellite based par on estimates of terrestrial net primary productivity’, *International Journal of Remote Sensing*, 31(19), pp. 5221–5237. doi: 10.1080/01431161.2010.496474.

R Core Team (2021) ‘R: A Language and Environment for Statistical Computing’. Vienna, Austria. Available at: https://www.r-project.org/.

Rembold, F. *et al.* (2017) ‘ASAP - Anomaly hot Spots of Agricultural Production, a new early warning decision support system developed by the Joint Research Centre’, *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images, MultiTemp 2017*. doi: 10.1109/Multi-Temp.2017.8035205.

Reichstein, M. *et al.* (2005) ‘On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review and improved algorithm’, *Global Change Biology*, 11(9), pp. 1424–1439. doi: 10.1111/j.1365-2486.2005.001002.x.

Richardson, A. D. *et al.* (2013) ‘Climate change, phenology, and phenological control of vegetation feedbacks to the climate system’, *Agricultural and Forest Meteorology*. Elsevier B.V., 169, pp. 156–173. doi: 10.1016/j.agrformet.2012.09.012.

Running, S., Mu, Q., Zhao, M. (2021) *MODIS/Aqua Gross Primary Productivity 8-Day L4 Global 500m SIN Grid V061, NASA EOSDIS Land Processes DAAC*. Available at: https://doi.org/10.5067/MODIS/MYD17A2H.061.

Rybchak, O. *et al.* (2020) ‘Multi-year CO2 budgets in South African semi-arid Karoo ecosystems under different grazing intensities’, *Biogeosciences Discussions*, (December), pp. 1–37.

SANBI (2018) ‘South African National Biodiversity Institute (2006- 2018)’, *The Vegetation Map of South Africa, Lesotho and Swaziland*. Available at: http://bgis.sanbi.org/SpatialDataset/Detail/18.

South African Government Department of Forestry and Fisheries (DFFE), South african National Land Use Classification Map, 2020 (https://egis.environment.gov.za/sa_national_land_cover_datasets), Accessed 15/11/2023

Stocker, B. D. *et al.* (2018) ‘Quantifying soil moisture impacts on light use efficiency across biomes’, *New Phytologist*, 218(4), pp. 1430–1449. doi: 10.1111/nph.15123.

Stocker, B. D. *et al.* (2020) ‘P-model v1.0: An optimality-based light use efficiency model for simulating ecosystem gross primary production’, *Geoscientific Model Development*, 13(3), pp. 1545–1581. doi: 10.5194/gmd-13-1545-2020.

Sulla-Menashe, D. and Friedl, M. A. (2018) ‘User Guide to Collection 6 MODIS Land Cover Dynamics (MCD12Q2) Product’, *User Guide*, 6(Figure 1), pp. 1–8. Available at: https://doi.org/10.5067/MODIS/MCD12Q1.006.

Thünen Institute (2022) ‘Eddy Covaraince Dataset’.

Thünen Institute of Climate-Smart Agriculture (2022) *Webpage of the Thünen Institute of Climate-Smart Agriculture*. Available at: https://www.thuenen.de/en/institutes/climate-smart-agriculture (Accessed: 11 October 2022).

Tian, F. *et al.* (2021) ‘Calibrating vegetation phenology from Sentinel-2 using eddy covariance, PhenoCam, and PEP725 networks across Europe’, *Remote Sensing of Environment*, 260(March). doi: 10.1016/j.rse.2021.112456.

Tobler, W. R. (1970) ‘A Computer Movie Simulating Urban Growth in the Detroit Region’, *Economic Geography*, 46, pp. 234–240.

VITO NV (2022) *Remote Sensing Virtual Machine*. Available at: https://remotesensing.vito.be/hubspot-topics/virtual-machine (Accessed: 26 September 2022).

Wang, X. *et al.* (2013) 'Validation of MODIS-GPP product at 10 flux sites in northern China', *International Journal of Remote Sensing*, 34(2), pp. 587–599. doi: 10.1080/01431161.2012.715774.

World Meteorological Organization (WMO) (2016) 'The Global Observing System For Climate Implementation Needs', *World Meteorological Organization*, 200(June), p. 316.

Wutzler, T. *et al.* (2018) 'Basic and extensible post-processing of eddy covariance flux data with REddyProc', *Biogeosciences*, 15(16), pp. 5015–5030. doi: 10.5194/bg-15-5015-2018.

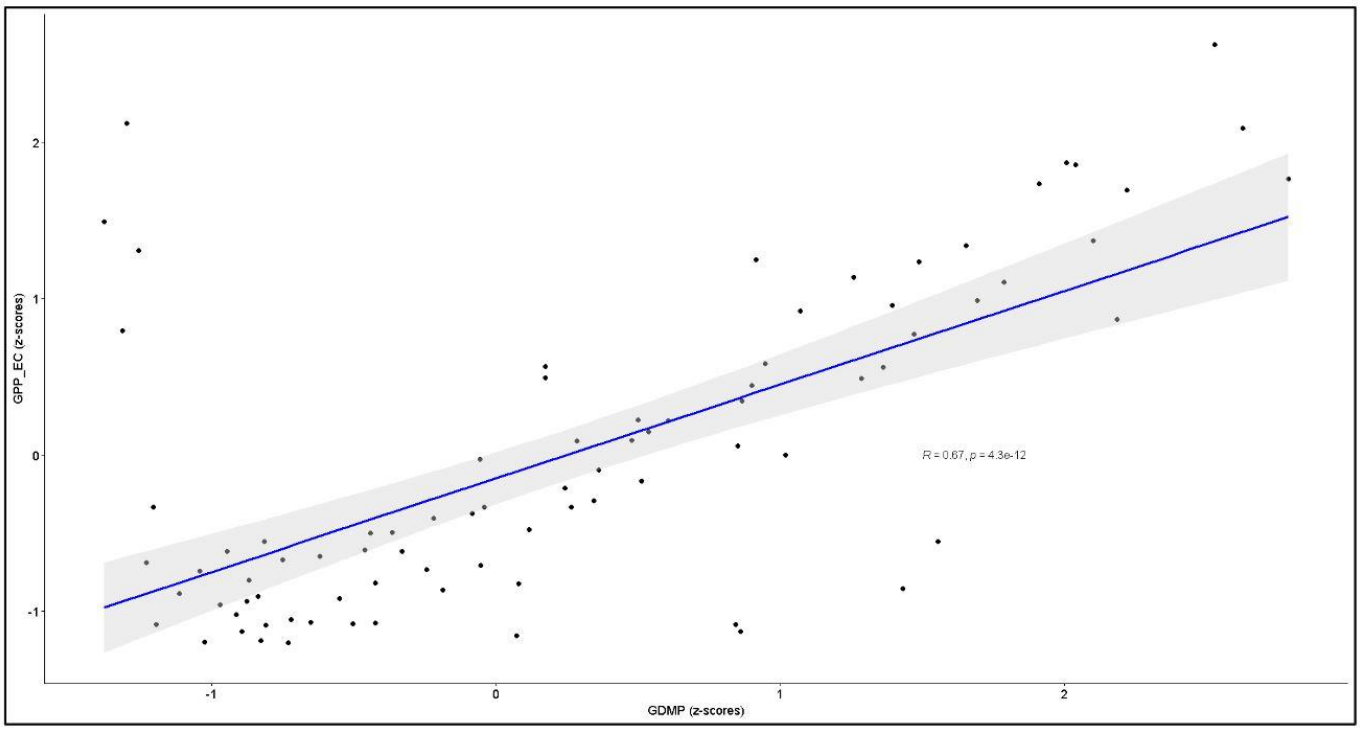
Zhang, F. *et al.* (2021) 'Integration of sentinel-3 olci land products and merra2 meteorology data into light use efficiency and vegetation index-driven models for modeling gross primary production', *Remote Sensing*, 13(5), pp. 1–23. doi: 10.3390/rs13051015.

Zhang, Z. *et al.* (2022) 'Improved estimation of global gross primary productivity during 1981–2020 using the optimized P model', *Science of the Total Environment*. Elsevier B.V., 838(March), p. 156172. doi: 10.1016/j.scitotenv.2022.156172.

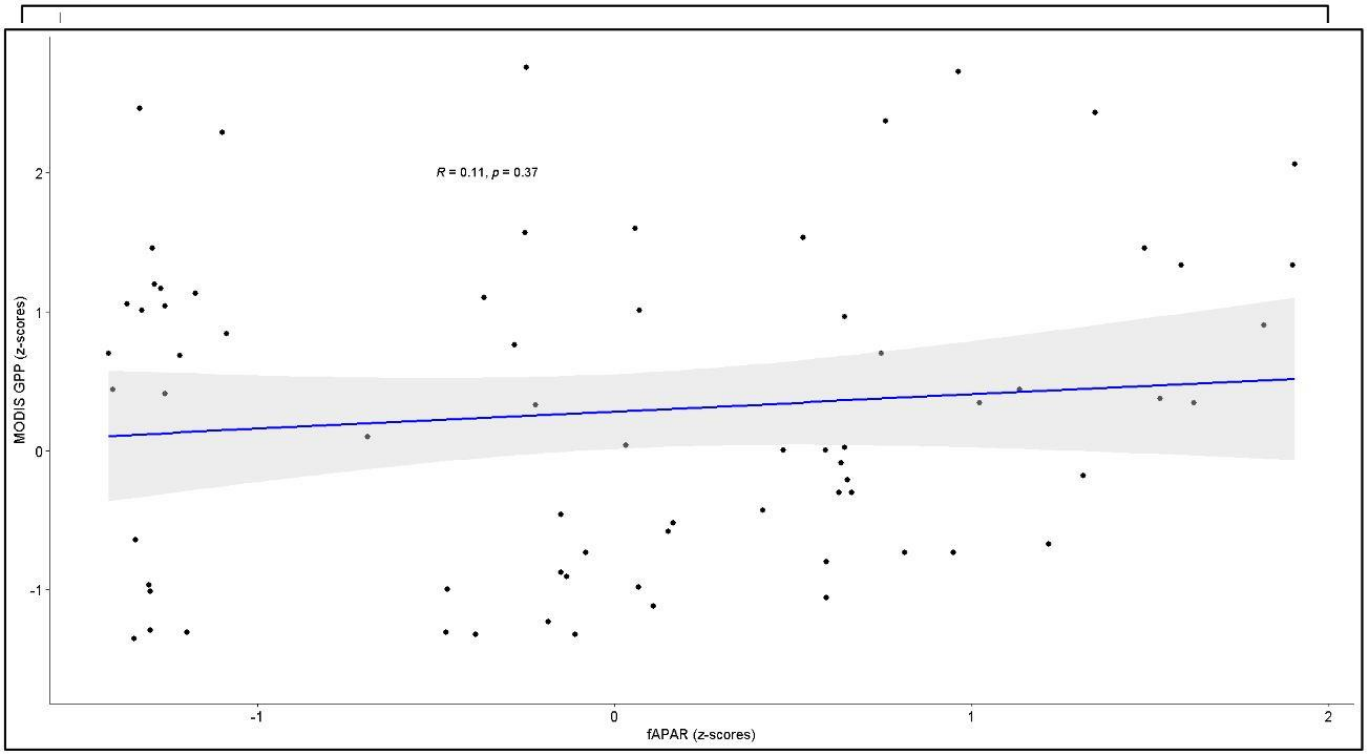
Zhang, Z., Zhao, L. and Lin, A. (2020) 'Evaluating the performance of Sentinel-3A OLCI land products for gross primary productivity estimation using AmeriFlux data', *Remote Sensing*, 12(12). doi: 10.3390/rs12121927.

8. Appendices

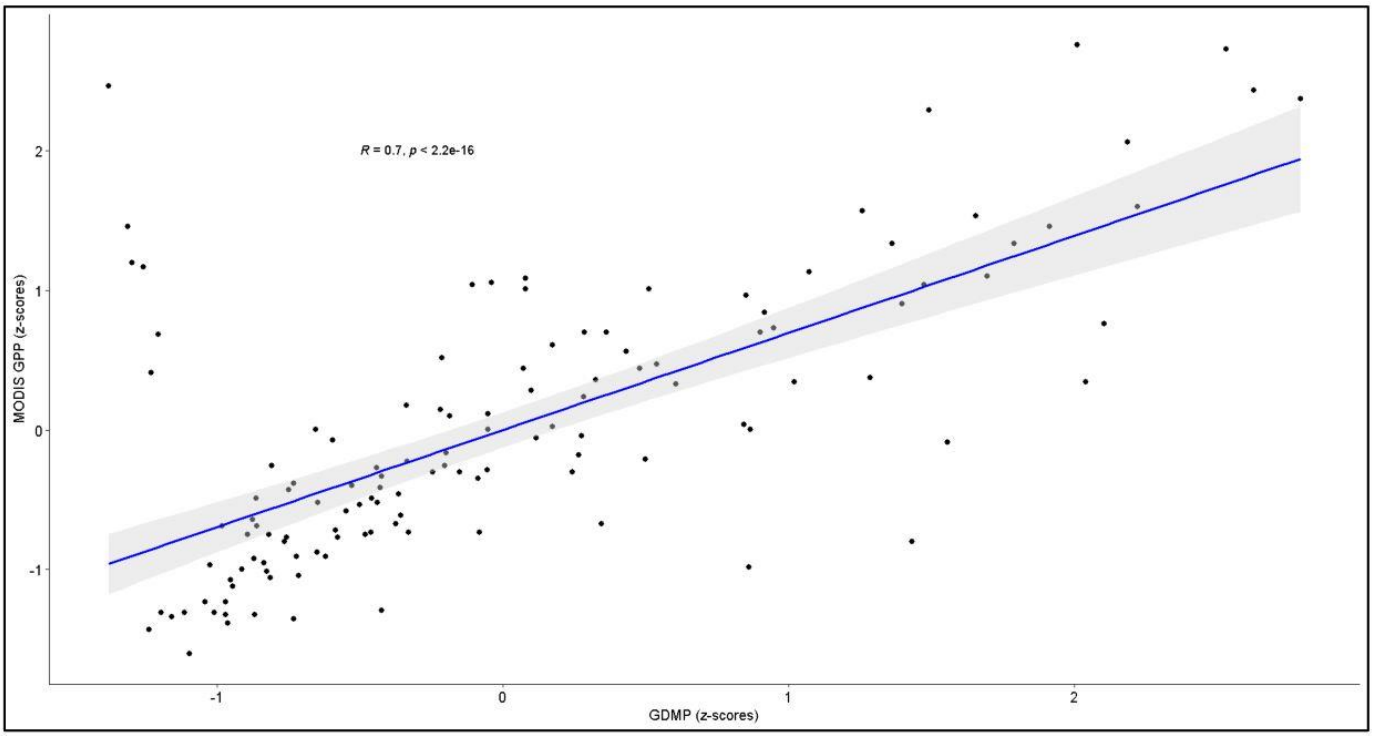
8.1 Appendix A – Scatterplots of Correlations



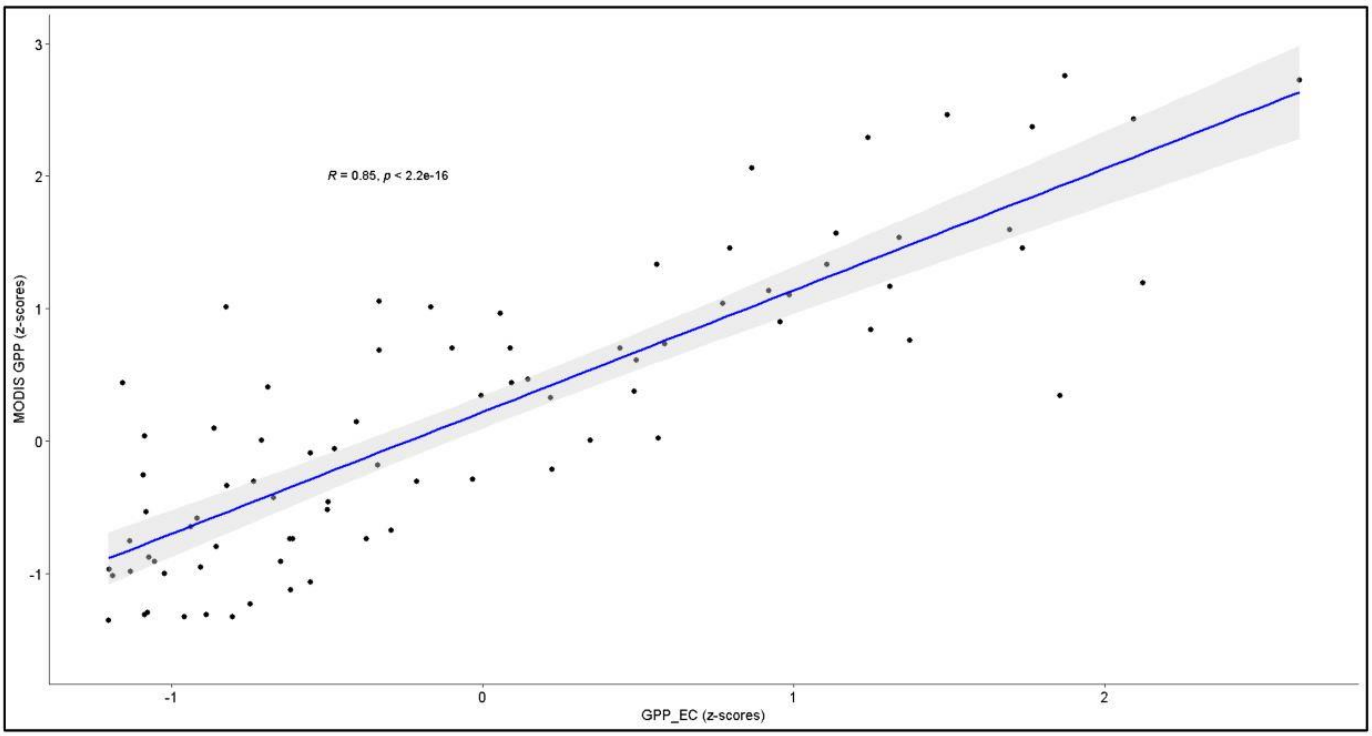
A.3. GDMP : GPP_{EC} (z-scores



A.4 - *f*APAR : $MODIS_{FPP}$ (z-scores)

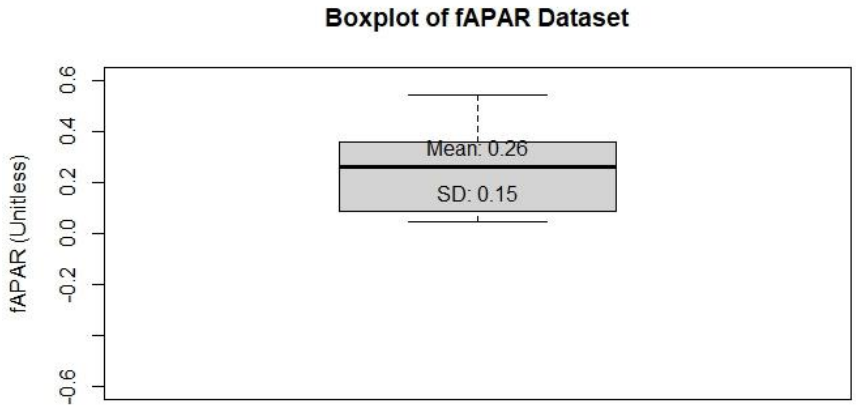


A.5 - GDMP : $MODIS_{GPP}$ (z-scores)

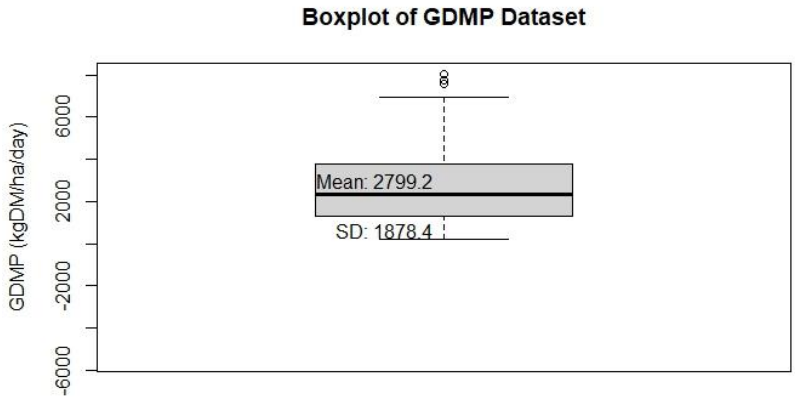


A.6 - GPP_{EC} : $MODIS_{GPP}$ (z-scores)

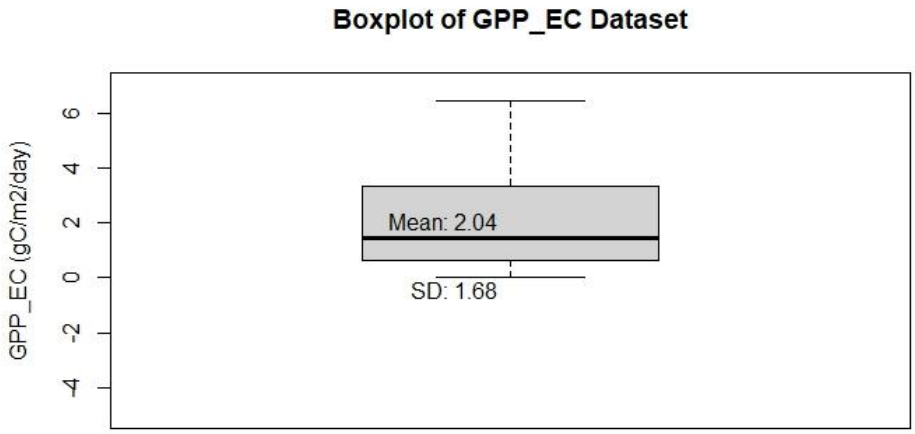
8.2 Appendix B - Boxplots of Datasets



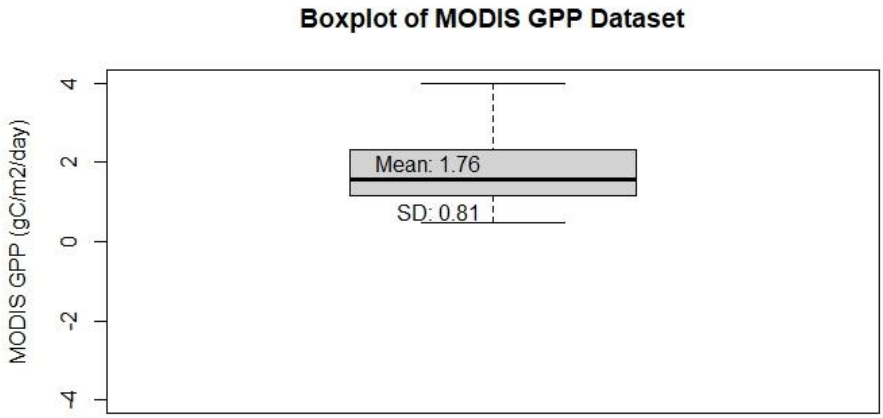
B.1 - Box Plot of *f*APAR Dataset



B.2 - Box Plot of GDMP Dataset



B.3 - Box Plot of Eddy Covariance GPP Dataset



B.4 - Box Plot of $MODIS_{GPP}$ Dataset

8.3 Appendix C - The Full Datasets

Record	Date	fAPAR	GDMP	mean_GPP_DT _micromolSec	mean_GPP_DT _gCm2day	MODIS_mean _kgCm2_8d	MODISGPP _gCm2day
1	2013/11/20	NA	1565.0	NA	NA	0.0141	1.7625
2	2013/11/30	NA	2397.1	NA	NA	0.0174	2.1750
3	2013/12/10	NA	2946.4	NA	NA	0.0211	2.6375
4	2013/12/20	NA	2986.6	NA	NA	0.0159	1.9875
5	2013/12/31	NA	2633.6	NA	NA	0.0118	1.4750
6	2014/01/10	NA	2164.9	NA	NA	0.0152	1.9000
7	2014/01/20	NA	2168.8	NA	NA	0.0126	1.5750
8	2014/01/31	NA	2595.4	NA	NA	0.0208	2.6000
9	2014/02/10	NA	3314.9	NA	NA	0.0138	1.7250
10	2014/02/20	NA	3407.0	NA	NA	0.0164	2.0500
11	2014/02/28	NA	3608.4	NA	NA	0.0177	2.2125
12	2014/03/10	NA	3330.6	NA	NA	0.0156	1.9500
13	2014/03/20	NA	2700.2	NA	NA	0.0148	1.8500
14	2014/03/31	NA	1989.7	NA	NA	0.0114	1.4250
15	2014/04/10	NA	1710.4	NA	NA	0.0091	1.1375
16	2014/04/20	NA	1455.0	NA	NA	0.0073	0.9125
17	2014/04/30	NA	973.2	NA	NA	0.0061	0.7625
18	2014/05/10	NA	902.9	NA	NA	0.0056	0.7000
19	2014/05/20	NA	619.6	NA	NA	0.0054	0.6750
20	2014/05/31	NA	736.8	NA	NA	0.0037	0.4625
21	2014/06/10	NA	411.6	NA	NA	0.0036	0.4500
22	2014/06/20	NA	360.4	NA	NA	0.0036	0.4500
23	2014/06/30	NA	392.1	NA	NA	0.0027	0.3375
24	2014/07/10	NA	314.2	NA	NA	0.0035	0.4375
25	2014/07/20	NA	320.2	NA	NA	0.0035	0.4375
26	2014/07/31	NA	347.7	NA	NA	0.0046	0.5750
27	2014/08/10	NA	396.1	NA	NA	0.0049	0.6125
28	2014/08/20	NA	485.4	NA	NA	0.0034	0.4250
29	2014/08/31	NA	390.2	NA	NA	0.0051	0.6375
30	2014/09/10	NA	733.0	NA	NA	0.0055	0.6875
31	2014/09/20	NA	667.2	NA	NA	0.0071	0.8875
32	2014/09/30	NA	644.8	NA	NA	0.0075	0.9375
33	2014/10/10	NA	681.0	NA	NA	0.0082	1.0250
34	2014/10/20	NA	742.8	NA	NA	0.0067	0.8375
35	2014/10/31	NA	950.4	NA	NA	0.0096	1.2000
36	2014/11/10	NA	1182.3	NA	NA	0.0096	1.2000
37	2014/11/20	NA	1424.1	NA	NA	0.0116	1.4500
38	2014/11/30	NA	1802.9	NA	NA	0.0115	1.4375
39	2014/12/10	NA	1681.3	NA	NA	0.0136	1.7000
40	2014/12/20	NA	1934.3	NA	NA	0.0109	1.3625
41	2014/12/31	NA	2126.9	NA	NA	0.0101	1.2625
42	2015/01/10	NA	1968.6	NA	NA	0.0123	1.5375
43	2015/01/20	NA	2426.4	NA	NA	0.0130	1.6250
44	2015/01/31	NA	1579.4	NA	NA	0.0107	1.3375
45	2015/02/10	NA	1699.2	NA	NA	0.0094	1.1750
46	2015/02/20	NA	1374.9	NA	NA	0.0091	1.1375
47	2015/02/28	NA	1260.4	NA	NA	0.0092	1.1500
48	2015/03/10	NA	1163.3	NA	NA	0.0081	1.0125
49	2015/03/20	NA	1177.8	NA	NA	0.0109	1.3625
50	2015/03/31	NA	1892.0	NA	NA	0.0092	1.1500
51	2015/04/10	NA	2414.2	NA	NA	0.0124	1.5500
52	2015/04/20	NA	2515.2	NA	NA	0.0121	1.5125
53	2015/04/30	NA	2093.0	NA	NA	0.0097	1.2125
54	2015/05/10	NA	1359.9	NA	NA	0.0089	1.1125
55	2015/05/20	NA	1007.7	NA	NA	0.0071	0.8875

Record	Date	fAPAR	GDMP	mean_GPP_DT _micromolSec	mean_GPP_DT _gCm2day	MODIS_mean _kgCm2_8d	MODISGPP _gCm2day
56	2015/05/31	NA	988.2	NA	NA	0.0051	0.6375
57	2015/06/10	NA	470.0	NA	NA	0.0048	0.6000
58	2015/06/20	NA	479.7	NA	NA	0.0045	0.5625
59	2015/06/30	NA	371.4	NA	NA	0.0049	0.6125
60	2015/07/10	NA	471.6	NA	NA	0.0033	0.4125
61	2015/07/20	NA	384.4	NA	NA	0.0026	0.3250
62	2015/07/31	NA	271.0	NA	NA	0.0037	0.4625
63	2015/08/10	NA	519.9	NA	NA	0.0061	0.7625
64	2015/08/20	NA	647.1	NA	NA	0.0062	0.7750
65	2015/08/31	NA	801.1	NA	NA	0.0068	0.8500
66	2015/09/10	NA	793.1	NA	NA	0.0079	0.9875
67	2015/09/20	NA	985.1	NA	NA	0.0105	1.3125
68	2015/09/30	NA	1462.2	NA	NA	0.0085	1.0625
69	2015/10/10	NA	1254.3	NA	NA	0.0081	1.0125
70	2015/10/20	NA	1101.0	NA	NA	0.0109	1.3625
71	2015/10/31	NA	902.6	NA	NA	0.0084	1.0500
72	2015/11/10	NA	932.1	0.0365	0.0379	0.0115	1.4375
73	2015/11/20	NA	1179.1	0.0204	0.0212	0.0079	0.9875
74	2015/11/30	NA	1009.7	0.0118	0.0123	0.0068	0.8500
75	2015/12/10	NA	694.8	0.0086	0.0090	0.0080	1.0000
76	2015/12/20	NA	863.0	0.0102	0.0106	0.0076	0.9500
77	2015/12/31	NA	713.3	0.0175	0.0181	0.0035	0.4375
78	2016/01/10	NA	734.7	0.1459	0.1514	0.0088	1.1000
79	2016/01/20	NA	1276.4	0.1985	0.2060	0.0124	1.5500
80	2016/01/31	NA	2385.3	1.3072	1.3564	0.0150	1.8750
81	2016/02/10	NA	2695.6	1.9158	1.9879	0.0122	1.5250
82	2016/02/20	NA	3476.1	1.8081	1.8762	0.0186	2.3250
83	2016/02/29	NA	4579.8	2.9148	3.0245	0.0188	2.3500
84	2016/03/10	NA	3124.9	2.7660	2.8702	0.0180	2.2500
85	2016/03/20	NA	3807.9	2.2029	2.2858	0.0171	2.1375
86	2016/03/31	NA	3018.1	1.1931	1.2380	0.0137	1.7125
87	2016/04/10	NA	2002.1	0.6350	0.6589	0.0119	1.4875
88	2016/04/20	NA	1854.3	0.2117	0.2197	0.0106	1.3250
89	2016/04/30	NA	1442.3	0.2575	0.2672	0.0082	1.0250
90	2016/05/10	NA	1118.4	0.1291	0.1339	0.0092	1.1500
91	2016/05/20	NA	1226.4	0.4958	0.5145	0.0079	0.9875
92	2016/05/31	NA	1286.7	0.7506	0.7789	0.0078	0.9750
93	2016/06/10	NA	1103.2	0.7260	0.7533	0.0060	0.7500
94	2016/06/20	NA	604.1	0.5380	0.5583	0.0059	0.7375
95	2016/06/30	NA	788.2	0.4795	0.4975	0.0051	0.6375
96	2016/07/10	NA	615.3	0.3705	0.3845	0.0054	0.6750
97	2016/07/20	NA	780.7	0.4987	0.5175	0.0048	0.6000
98	2016/07/31	NA	383.6	0.3979	0.4129	0.0054	0.6750
99	2016/08/10	NA	728.3	0.4980	0.5168	0.0048	0.6000
100	2016/08/20	31.67	770.7	0.3417	0.3546	0.0069	0.8625
101	2016/08/31	28.67	1277.7	0.0619	0.0643	0.0090	1.1250
102	2016/09/10	27.89	1345.9	0.0592	0.0615	0.0086	1.0750
103	2016/09/20	28.33	1122.2	0.0407	0.0422	0.0092	1.1500
104	2016/09/30	28.56	1153.8	0.0168	0.0175	0.0075	0.9375
105	2016/10/10	25.22	1007.8	0.0621	0.0644	0.0081	1.0125
106	2016/10/20	20.33	945.2	0.1793	0.1860	0.0089	1.1125
107	2016/10/31	18.44	1060.4	0.1975	0.2049	0.0080	1.0000
108	2016/11/10	16.56	905.0	0.0649	0.0674	0.0095	1.1875
109	2016/11/20	15.44	1169.8	0.0290	0.0301	0.0102	1.2750
110	2016/11/30	14.22	1111.0	0.0856	0.0888	0.0070	0.8750
111	2016/12/10	15.00	1048.2	0.2179	0.2261	0.0092	1.1500
112	2016/12/20	16.33	770.8	0.0178	0.0184	0.0061	0.7625

Record	Date	fAPAR	GDMP	mean_GPP_DT _micromolSec	mean_GPP_DT _gCm2day	MODIS_mean _kgCm2_8d	MODISGPP _gCm2day
113	2016/12/31	15.44	1999.2	0.2188	0.2270	0.0057	0.7125
114	2017/01/10	13.00	2721.6	1.4262	1.4799	0.0209	2.6125
115	2017/01/20	23.44	4515.9	3.9881	4.1383	0.0195	2.4375
116	2017/01/31	56.00	3936.4	2.3214	2.4088	0.0162	2.0250
117	2017/02/10	66.89	3759.6	1.6959	1.7598	0.0206	2.5750
118	2017/02/20	54.89	5163.1	3.8084	3.9518	0.0242	3.0250
119	2017/02/28	66.44	6967.4	4.7123	4.8898	0.0244	3.0500
120	2017/03/10	102.67	6628.8	4.9741	5.1614	0.0163	2.0375
121	2017/03/20	121.67	5214.7	2.7556	2.8594	0.0165	2.0625
122	2017/03/31	109.89	3446.7	1.4899	1.5461	0.0097	1.2125
123	2017/04/10	87.89	2338.2	0.7719	0.8009	0.0121	1.5125
124	2017/04/20	69.89	1765.4	0.4770	0.4950	0.0103	1.2875
125	2017/04/30	58.67	1575.8	0.2260	0.2345	0.0084	1.0500
126	2017/05/10	46.67	1083.4	0.3080	0.3196	0.0076	0.9500
127	2017/05/20	35.78	811.0	0.4673	0.4849	0.0068	0.8500
128	2017/05/31	32.67	682.4	0.2075	0.2153	0.0050	0.6250
129	2017/06/10	30.78	576.7	0.1069	0.1109	0.0044	0.5500
130	2017/06/20	25.33	482.0	0.0535	0.0555	0.0039	0.4875
131	2017/06/30	23.78	373.4	0.0455	0.0472	0.0043	0.5375
132	2017/07/10	22.78	402.6	0.0768	0.0797	0.0029	0.3625
133	2017/07/20	20.00	297.0	0.1212	0.1257	0.0040	0.5000
134	2017/07/31	18.22	393.8	0.1634	0.1696	0.0054	0.6750
135	2017/08/10	16.33	442.7	0.1951	0.2024	0.0039	0.4875
136	2017/08/20	14.67	404.4	0.2055	0.2133	0.0036	0.4500
137	2017/08/31	14.67	472.9	0.2398	0.2488	0.0064	0.8000
138	2017/09/10	15.67	738.8	0.3141	0.3259	0.0065	0.8125
139	2017/09/20	14.56	643.6	0.3515	0.3647	0.0065	0.8125
140	2017/09/30	14.22	690.8	0.3687	0.3826	0.0072	0.9000
141	2017/10/10	13.67	537.3	0.3768	0.3910	0.0085	1.0625
142	2017/10/20	13.56	984.9	0.5502	0.5709	0.0101	1.2625
143	2017/10/31	13.89	1107.2	0.9150	0.9495	0.0083	1.0375
144	2017/11/10	16.44	1029.9	0.2956	0.3067	0.0083	1.0375
145	2017/11/20	16.89	976.4	0.3846	0.3991	0.0098	1.2250
146	2017/11/30	16.44	1233.6	0.5449	0.5655	0.0088	1.1000
147	2017/12/10	17.11	975.1	0.0954	0.0990	0.0094	1.1750
148	2017/12/20	15.56	1079.3	0.0662	0.0687	0.0099	1.2375
149	2017/12/31	14.22	980.4	0.0378	0.0392	0.0056	0.7000
150	2018/01/10	14.33	717.3	0.0823	0.0854	0.0077	0.9625
151	2018/01/20	13.89	1150.4	0.4458	0.4625	0.0099	1.2375
152	2018/01/31	11.11	3332.2	2.1090	2.1884	0.0186	2.3250
153	2018/02/10	23.00	5597.1	3.9724	4.1220	0.0289	3.6125
154	2018/02/20	55.00	6572.6	4.9984	5.1867	0.0319	3.9875
155	2018/02/28	92.78	8039.6	4.8297	5.0116	0.0294	3.6750
156	2018/03/10	120.00	6389.4	4.7784	4.9584	0.0235	2.9375
157	2018/03/20	132.56	5421.0	3.5144	3.6467	0.0199	2.4875
158	2018/03/31	125.22	4711.9	1.9596	2.0334	0.0163	2.0375
159	2018/04/10	106.89	3695.4	2.1175	2.1973	0.0169	2.1125
160	2018/04/20	88.44	3125.4	2.8825	2.9911	0.0142	1.7750
161	2018/04/30	88.78	3736.6	2.3289	2.4166	0.0127	1.5875
162	2018/05/10	94.78	2644.0	1.3595	1.4107	0.0093	1.1625
163	2018/05/20	86.56	1266.7	1.0684	1.1086	0.0072	0.9000
164	2018/05/31	68.33	1022.7	0.9654	1.0017	0.0068	0.8500
165	2018/06/10	57.33	839.9	0.7564	0.7849	0.0061	0.7625
166	2018/06/20	46.56	705.2	0.5270	0.5468	0.0056	0.7000
167	2018/06/30	39.56	687.6	0.3605	0.3741	0.0051	0.6375
168	2018/07/10	35.11	485.8	0.2265	0.2350	0.0039	0.4875
169	2018/07/20	32.89	449.6	0.2365	0.2454	0.0056	0.7000

Record	Date	fAPAR	GDMP	mean_GPP_DT _micromolSec	mean_GPP_DT _gCm2day	MODIS_mean _kgCm2_8d	MODISGPP _gCm2day
170	2018/07/31	28.89	621.2	0.2230	0.2314	0.0055	0.6875
171	2018/08/10	24.67	389.4	0.1860	0.1930	0.0042	0.5250
172	2018/08/20	22.56	672.1	0.0629	0.0653	0.0052	0.6500
173	2018/08/31	23.22	687.8	0.0815	0.0846	0.0055	0.6875
174	2018/09/10	22.78	473.9	0.2574	0.2671	0.0068	0.8500
175	2018/09/20	20.56	853.9	0.4579	0.4752	0.0077	0.9625
176	2018/09/30	18.00	1075.3	0.6176	0.6409	0.0082	1.0250
177	2018/10/10	18.22	1074.2	0.4439	0.4606	0.0096	1.2000
178	2018/10/20	18.11	958.2	0.4076	0.4229	0.0080	1.0000
179	2018/10/31	18.89	1084.8	0.4127	0.4282	0.0091	1.1375
180	2018/11/10	18.56	1182.3	0.0694	0.0720	0.0065	0.8125
181	2018/11/20	17.22	944.1	0.0183	0.0190	0.0082	1.0250
182	2018/11/30	17.00	1069.6	0.0092	0.0096	0.0095	1.1875
183	2018/12/10	16.67	1116.3	0.0475	0.0493	0.0069	0.8625
184	2018/12/20	16.00	1003.7	0.0208	0.0216	0.0074	0.9250
185	2018/12/31	15.22	905.0	0.0202	0.0210	0.0048	0.6000
186	2019/01/10	15.67	972.4	0.0248	0.0257	0.0077	0.9625
187	2019/01/20	15.33	874.2	0.0200	0.0208	0.0078	0.9750
188	2019/01/31	15.44	1245.9	0.0384	0.0399	0.0075	0.9375
189	2019/02/10	14.56	2947.8	0.6283	0.6520	0.0206	2.5750
190	2019/02/20	17.00	5565.0	3.2155	3.3366	0.0208	2.6000
191	2019/02/28	50.56	5982.2	3.5642	3.6985	0.0212	2.6500
192	2019/03/10	88.44	4396.1	2.0594	2.1369	0.0203	2.5375
193	2019/03/20	92.33	4489.1	2.6832	2.7843	0.0186	2.3250
194	2019/03/31	82.00	4425.9	2.5270	2.6222	0.0141	1.7625
195	2019/04/10	89.22	3256.4	1.6217	1.6828	0.0121	1.5125
196	2019/04/20	86.44	2698.4	0.8154	0.8461	0.0141	1.7625
197	2019/04/30	70.44	1970.9	1.1565	1.2001	0.0107	1.3375
198	2019/05/10	58.67	2113.8	1.1603	1.2040	0.0111	1.3875
199	2019/05/20	61.22	1930.1	0.9766	1.0134	0.0093	1.1625
200	2019/05/31	60.11	975.0	0.4127	0.4283	0.0055	0.6875
201	2019/06/10	50.56	681.8	0.2852	0.2959	0.0045	0.5625
202	2019/06/20	36.44	624.3	0.1271	0.1319	0.0045	0.5625
203	2019/06/30	28.67	509.9	0.1469	0.1525	0.0046	0.5750
204	2019/07/10	26.56	450.7	0.1106	0.1148	0.0049	0.6125
205	2019/07/20	24.33	546.7	0.0946	0.0982	0.0039	0.4875
206	2019/07/31	21.78	492.1	0.0421	0.0437	0.0048	0.6000
207	2019/08/10	19.67	521.9	0.0261	0.0271	0.0056	0.7000
208	2019/08/20	17.89	585.4	0.0242	0.0251	0.0055	0.6875
209	2019/08/31	17.11	648.1	0.0085	0.0088	0.0054	0.6750
210	2019/09/10	16.33	780.7	0.0066	0.0068	0.0049	0.6125
211	2019/09/20	15.67	769.8	0.0166	0.0172	0.0063	0.7875
212	2019/09/30	15.56	755.3	0.0353	0.0366	0.0083	1.0375
213	2019/10/10	15.33	892.1	0.0521	0.0541	0.0074	0.9250
214	2019/10/20	15.33	943.2	0.0594	0.0617	0.0086	1.0750
215	2019/10/31	15.00	961.2	0.0295	0.0306	0.0080	1.0000
216	2019/11/10	14.89	988.2	0.0135	0.0140	0.0096	1.2000

8.4 Appendix D - The Interim Analysis Dataset

Record	Date	fAPAR	GDMP	mean_GPP_DT _micromolSec	mean_GPP _DT_gCm2day	MODIS_mean _kgCm2_8d	MODISGPP _gCm2day
1	2013/11/20	NA	1565	NA	NA	0.0141	1.7625
2	2013/11/30	NA	2397.11	NA	NA	0.0174	2.175
3	2013/12/10	NA	2946.44	NA	NA	0.0211	2.6375
4	2013/12/20	NA	2986.56	NA	NA	0.0159	1.9875
5	2013/12/31	NA	2633.56	NA	NA	0.0118	1.475
6	2014/01/10	NA	2164.89	NA	NA	0.0152	1.9
7	2014/01/20	NA	2168.78	NA	NA	0.0126	1.575
8	2014/01/31	NA	2595.44	NA	NA	0.0208	2.6
9	2014/02/10	NA	3314.89	NA	NA	0.0138	1.725
10	2014/02/20	NA	3407	NA	NA	0.0164	2.05
11	2014/02/28	NA	3608.44	NA	NA	0.0177	2.2125
12	2014/03/10	NA	3330.56	NA	NA	0.0156	1.95
13	2014/03/20	NA	2700.22	NA	NA	0.0148	1.85
14	2014/03/31	NA	1989.67	NA	NA	0.0114	1.425
15	2014/04/10	NA	1710.44	NA	NA	0.0091	1.1375
16	2014/04/20	NA	1455	NA	NA	0.0073	0.9125
17	2014/04/30	NA	973.22	NA	NA	0.0061	0.7625
18	2014/05/10	NA	902.89	NA	NA	0.0056	0.7
19	2014/05/20	NA	619.56	NA	NA	0.0054	0.675
20	2014/05/31	NA	736.78	NA	NA	0.0037	0.4625
21	2014/06/10	NA	NA	NA	NA	NA	NA
22	2014/07/10	NA	NA	NA	NA	NA	NA
23	2014/08/10	NA	NA	NA	NA	NA	NA
24	2014/09/10	NA	NA	NA	NA	NA	NA
35	2014/10/31	NA	950.44	NA	NA	0.0096	1.2
36	2014/11/10	NA	1182.33	NA	NA	0.0096	1.2
37	2014/11/20	NA	1424.11	NA	NA	0.0116	1.45
38	2014/11/30	NA	1802.89	NA	NA	0.0115	1.4375
39	2014/12/10	NA	1681.33	NA	NA	0.0136	1.7
40	2014/12/20	NA	1934.33	NA	NA	0.0109	1.3625
41	2014/12/31	NA	2126.89	NA	NA	0.0101	1.2625
42	2015/01/10	NA	1968.56	NA	NA	0.0123	1.5375
43	2015/01/20	NA	2426.44	NA	NA	0.013	1.625
44	2015/01/31	NA	1579.44	NA	NA	0.0107	1.3375
45	2015/02/10	NA	1699.22	NA	NA	0.0094	1.175
46	2015/02/20	NA	1374.89	NA	NA	0.0091	1.1375
47	2015/02/28	NA	1260.44	NA	NA	0.0092	1.15
48	2015/03/10	NA	1163.33	NA	NA	0.0081	1.0125
49	2015/03/20	NA	1177.78	NA	NA	0.0109	1.3625
50	2015/03/31	NA	1892	NA	NA	0.0092	1.15
51	2015/04/10	NA	2414.22	NA	NA	0.0124	1.55
52	2015/04/20	NA	2515.22	NA	NA	0.0121	1.5125
53	2015/04/30	NA	2093	NA	NA	0.0097	1.2125

54	2015/05/10	NA	1359.89	NA	NA	0.0089	1.1125
55	2015/05/20	NA	1007.67	NA	NA	0.0071	0.8875
56	2015/05/31	NA	988.22	NA	NA	0.0051	0.6375
57	2015/06/10	NA	470	NA	NA	0.0048	0.6
58	2015/07/10	NA	NA	NA	NA	NA	NA
59	2015/08/10	NA	NA	NA	NA	NA	NA
60	2015/09/10	NA	NA	NA	NA	NA	NA
61	2015/10/10	NA	NA	NA	NA	NA	NA
62	2015/11/10	NA	NA	NA	NA	NA	NA
63	2015/12/10	NA	NA	NA	NA	NA	NA
79	2016/01/20	NA	1276.44	0.1985	0.206	0.0124	1.55
80	2016/01/31	NA	2385.33	1.3072	1.3564	0.015	1.875
81	2016/02/10	NA	2695.56	1.9158	1.9879	0.0122	1.525
82	2016/02/20	NA	3476.11	1.8081	1.8762	0.0186	2.325
83	2016/02/29	NA	4579.78	2.9148	3.0245	0.0188	2.35
84	2016/03/10	NA	3124.89	2.766	2.8702	0.018	2.25
85	2016/03/20	NA	3807.89	2.2029	2.2858	0.0171	2.1375
86	2016/03/31	NA	3018.11	1.1931	1.238	0.0137	1.7125
87	2016/04/10	NA	2002.11	0.635	0.6589	0.0119	1.4875
88	2016/04/20	NA	1854.33	0.2117	0.2197	0.0106	1.325
89	2016/04/30	NA	1442.33	0.2575	0.2672	0.0082	1.025
90	2016/05/10	NA	1118.44	0.1291	0.1339	0.0092	1.15
91	2016/05/20	NA	1226.44	0.4958	0.5145	0.0079	0.9875
92	2016/06/10	NA	NA	NA	NA	NA	NA
93	2016/07/10	NA	NA	NA	NA	NA	NA
94	2016/08/10	NA	NA	NA	NA	NA	NA
95	2016/09/10	NA	NA	NA	NA	NA	NA
96	2016/10/10	NA	NA	NA	NA	NA	NA
97	2016/11/10	NA	NA	NA	NA	NA	NA
113	2016/12/31	15.44	1999.22	0.2188	0.227	0.0057	0.7125
114	2017/01/10	13.00	2721.56	1.4262	1.4799	0.0209	2.6125
115	2017/01/20	23.44	4515.89	3.9881	4.1383	0.0195	2.4375
116	2017/01/31	56.00	3936.44	2.3214	2.4088	0.0162	2.025
117	2017/02/10	66.89	3759.56	1.6959	1.7598	0.0206	2.575
118	2017/02/20	54.89	5163.11	3.8084	3.9518	0.0242	3.025
119	2017/02/28	66.44	6967.44	4.7123	4.8898	0.0244	3.05
120	2017/03/10	102.67	6628.78	4.9741	5.1614	0.0163	2.0375
121	2017/03/20	121.67	5214.67	2.7556	2.8594	0.0165	2.0625
122	2017/03/31	109.89	3446.67	1.4899	1.5461	0.0097	1.2125
123	2017/04/10	87.89	2338.22	0.7719	0.8009	0.0121	1.5125
124	2017/04/20	69.89	1765.44	0.477	0.495	0.0103	1.2875
125	2017/04/30	58.67	1575.78	0.226	0.2345	0.0084	1.05
126	2017/05/10	46.67	1083.44	0.308	0.3196	0.0076	0.95
127	2017/06/10	NA	NA	NA	NA	NA	NA
128	2017/07/10	NA	NA	NA	NA	NA	NA
129	2017/08/10	NA	NA	NA	NA	NA	NA

130	2017/09/10	NA	NA	NA	NA	NA	NA
131	2017/10/10	NA	NA	NA	NA	NA	NA
132	2017/11/10	NA	NA	NA	NA	NA	NA
133	2017/12/10	NA	NA	NA	NA	NA	NA
151	2018/01/20	13.89	1150.44	0.4458	0.4625	0.0099	1.2375
152	2018/01/31	11.11	3332.22	2.109	2.1884	0.0186	2.325
153	2018/02/10	23.00	5597.11	3.9724	4.122	0.0289	3.6125
154	2018/02/20	55.00	6572.56	4.9984	5.1867	0.0319	3.9875
155	2018/02/28	92.78	8039.56	4.8297	5.0116	0.0294	3.675
156	2018/03/10	120.00	6389.44	4.7784	4.9584	0.0235	2.9375
157	2018/03/20	132.56	5421	3.5144	3.6467	0.0199	2.4875
158	2018/03/31	125.22	4711.89	1.9596	2.0334	0.0163	2.0375
159	2018/04/10	106.89	3695.44	2.1175	2.1973	0.0169	2.1125
160	2018/04/20	88.44	3125.44	2.8825	2.9911	0.0142	1.775
161	2018/04/30	88.78	3736.56	2.3289	2.4166	0.0127	1.5875
162	2018/05/10	94.78	2644	1.3595	1.4107	0.0093	1.1625
163	2018/05/20	86.56	1266.67	1.0684	1.1086	0.0072	0.9
164	2018/05/31	68.33	1022.67	0.9654	1.0017	0.0068	0.85
165	2018/06/10	57.33	839.89	0.7564	0.7849	0.0061	0.7625
166	2018/06/20	46.56	705.22	0.527	0.5468	0.0056	0.7
167	2018/07/20	NA	NA	NA	NA	NA	NA
168	2018/08/20	NA	NA	NA	NA	NA	NA
169	2018/09/20	NA	NA	NA	NA	NA	NA
170	2018/10/20	NA	NA	NA	NA	NA	NA
171	2018/11/20	NA	NA	NA	NA	NA	NA
172	2018/12/20	NA	NA	NA	NA	NA	NA
187	2019/01/20	15.33	874.22	0.02	0.0208	0.0078	0.975
188	2019/01/31	15.44	1245.89	0.0384	0.0399	0.0075	0.9375
189	2019/02/10	14.56	2947.78	0.6283	0.652	0.0206	2.575
190	2019/02/20	17.00	5565	3.2155	3.3366	0.0208	2.6
191	2019/02/28	50.56	5982.22	3.5642	3.6985	0.0212	2.65
192	2019/03/10	88.44	4396.11	2.0594	2.1369	0.0203	2.5375
193	2019/03/20	92.33	4489.11	2.6832	2.7843	0.0186	2.325
194	2019/03/31	82.00	4425.89	2.527	2.6222	0.0141	1.7625
195	2019/04/10	89.22	3256.44	1.6217	1.6828	0.0121	1.5125
196	2019/04/20	86.44	2698.44	0.8154	0.8461	0.0141	1.7625
197	2019/04/30	70.44	1970.89	1.1565	1.2001	0.0107	1.3375
198	2019/05/10	58.67	2113.78	1.1603	1.204	0.0111	1.3875
199	2019/05/20	61.22	1930.11	0.9766	1.0134	0.0093	1.1625
200	2019/05/31	60.11	975	0.4127	0.4283	0.0055	0.6875
201	2019/06/10	NA	NA	NA	NA	NA	NA
202	2019/07/10	NA	NA	NA	NA	NA	NA
203	2019/08/10	NA	NA	NA	NA	NA	NA
204	2019/09/10	NA	NA	NA	NA	NA	NA
205	2019/10/10	NA	NA	NA	NA	NA	NA
206	2019/11/10	NA	NA	NA	NA	NA	NA

221	2019/12/31	13.78	1423.67	0.0184	0.0191	0.0053	0.6625
222	2020/01/10	11.56	2935.33	0.0912	0.0947	0.0169	2.1125
223	2020/01/20	20.22	4812.11	3.4557	3.5859	0.0214	2.675
224	2020/01/31	53.78	6748.78	4.1913	4.3491	0.019	2.375
225	2020/02/10	84.11	5906.89	4.1352	4.291	0.024	3
226	2020/02/20	100.44	7550.22	6.2194	6.4536	0.0317	3.9625
227	2020/02/29	114.78	7735.22	5.3557	5.5574	0.0298	3.725
228	2020/03/10	135.78	6904.67	3.3682	3.4951	0.0274	3.425
229	2020/03/20	135.56	6158.33	3.7585	3.9001	0.0227	2.8375
230	2020/03/31	123.89	5357.89	2.8751	2.9834	0.0227	2.8375
231	**2020/04/10**	**113.56(DN) * 0.004 = 0.454**	3296.78	1.4203	1.4738	0.0129	1.6125
232	2020/04/20	99.89	2177.67	0.962	0.9982	0.0093	1.1625
233	2020/04/30	79.89	1388.22	0.878	0.911	0.0113	1.4125
234	2020/05/10	59.22	1634.78	0.9133	0.9476	0.0082	1.025
235	2020/05/20	49.67	1165.67	0.6614	0.6863	0.0055	0.6875
236	2020/06/10	NA	NA	NA	NA	NA	NA
237	2020/07/10	NA	NA	NA	NA	NA	NA
238	2020/08/10	NA	NA	NA	NA	NA	NA
239	2020/09/10	NA	NA	NA	NA	NA	NA
240	2020/10/10	NA	NA	NA	NA	NA	NA
241	2020/11/10	NA	NA	NA	NA	NA	NA
257	2020/12/31	19.33	553.11	0.2045	0.2122	0.0056	0.7
258	2021/01/10	17.00	486.44	0.8479	0.8798	0.0167	2.0875
259	2021/01/20	16.56	435.78	4.0891	4.2431	0.0216	2.7
260	2021/01/31	14.33	208.11	4.3874	4.5526	0.03	3.75
261	2021/02/10	15.89	359.67	5.4039	5.6074	0.0218	2.725
262	2021/02/20	15.67	330.78	3.2555	3.3781	0.0235	2.9375
263	2021/02/28	18.56	532	1.4261	1.4798	0.0185	2.3125
264	2021/03/10	38.33	2447.22	0.5658	0.5871	0.0147	1.8375
265	2021/03/20	65.44	4381.67	0.205	0.2127	0.0143	1.7875
266	2021/03/31	88.11	5720	1.0679	1.1081	0.0135	1.6875
267	2021/04/10	86.56	5486.89	0.5795	0.6013	0.0089	1.1125
268	2021/04/20	66.78	4416.11	0.1306	0.1355	0.0077	0.9625

8.5 Appendix E - The Anomaly Analysis Dataset (z-scores)

Record	Date	fAPAR_z	GDMP_z	GPP_EC_z	MODIS_GPP_z
1	2013/11/20	NA	-0.6570	NA	0.0051
2	2013/11/30	NA	-0.2141	NA	0.5155
3	2013/12/10	NA	0.0784	NA	1.0877
4	2013/12/20	NA	0.0997	NA	0.2835
5	2013/12/31	NA	-0.0882	NA	-0.3506
6	2014/01/10	NA	-0.3377	NA	0.1752
7	2014/01/20	NA	-0.3356	NA	-0.2269
8	2014/01/31	NA	-0.1085	NA	1.0413
9	2014/02/10	NA	0.2745	NA	-0.0413
10	2014/02/20	NA	0.3236	NA	0.3608
11	2014/02/28	NA	0.4308	NA	0.5619
12	2014/03/10	NA	0.2829	NA	0.2371
13	2014/03/20	NA	-0.0527	NA	0.1134
14	2014/03/31	NA	-0.4310	NA	-0.4125
15	2014/04/10	NA	-0.5796	NA	-0.7682
16	2014/04/20	NA	-0.7156	NA	-1.0465
17	2014/04/30	NA	-0.9721	NA	-1.2321
18	2014/05/10	NA	-1.0095	NA	-1.3094
19	2014/05/20	NA	-1.1604	NA	-1.3404
20	2014/05/31	NA	-1.0980	NA	-1.6033
21	2014/06/10	NA	NA	NA	NA
22	2014/07/10	NA	NA	NA	NA
23	2014/08/10	NA	NA	NA	NA
24	2014/09/10	NA	NA	NA	NA
35	2014/10/31	NA	-0.9842	NA	-0.6908
36	2014/11/10	NA	-0.8608	NA	-0.6908
37	2014/11/20	NA	-0.7321	NA	-0.3815
38	2014/11/30	NA	-0.5304	NA	-0.3970
39	2014/12/10	NA	-0.5951	NA	-0.0722
40	2014/12/20	NA	-0.4604	NA	-0.4898
41	2014/12/31	NA	-0.3579	NA	-0.6135
42	2015/01/10	NA	-0.4422	NA	-0.2733
43	2015/01/20	NA	-0.1984	NA	-0.1650
44	2015/01/31	NA	-0.6494	NA	-0.5207
45	2015/02/10	NA	-0.5856	NA	-0.7218
46	2015/02/20	NA	-0.7583	NA	-0.7682
47	2015/02/28	NA	-0.8192	NA	-0.7527
48	2015/03/10	NA	-0.8709	NA	-0.9228
49	2015/03/20	NA	-0.8632	NA	-0.4898
50	2015/03/31	NA	-0.4830	NA	-0.7527
51	2015/04/10	NA	-0.2049	NA	-0.2578
52	2015/04/20	NA	-0.1512	NA	-0.3042
53	2015/04/30	NA	-0.3760	NA	-0.6754

54	2015/05/10	NA	-0.7662	NA	-0.7991
55	2015/05/20	NA	-0.9538	NA	-1.0775
56	2015/05/31	NA	-0.9641	NA	-1.3868
57	2015/06/10	NA	-1.2400	NA	-1.4332
58	2015/07/10	NA	NA	NA	NA
59	2015/08/10	NA	NA	NA	NA
60	2015/09/10	NA	NA	NA	NA
61	2015/10/10	NA	NA	NA	NA
62	2015/11/10	NA	NA	NA	NA
63	2015/12/10	NA	NA	NA	NA
79	2016/01/20	NA	-0.8107	-1.0912	-0.2578
80	2016/01/31	NA	-0.2203	-0.4068	0.1443
81	2016/02/10	NA	-0.0552	-0.0312	-0.2887
82	2016/02/20	NA	0.3604	-0.0976	0.7011
83	2016/02/29	NA	0.9479	0.5854	0.7320
84	2016/03/10	NA	0.1734	0.4936	0.6083
85	2016/03/20	NA	0.5370	0.1460	0.4691
86	2016/03/31	NA	0.1165	-0.4773	-0.0567
87	2016/04/10	NA	-0.4243	-0.8218	-0.3351
88	2016/04/20	NA	-0.5030	-1.0830	-0.5362
89	2016/04/30	NA	-0.7224	-1.0548	-0.9073
90	2016/05/10	NA	-0.8948	-1.1340	-0.7527
91	2016/05/20	NA	-0.8373	-0.9076	-0.9537
92	2016/06/10	NA	NA	NA	NA
93	2016/07/10	NA	NA	NA	NA
94	2016/08/10	NA	NA	NA	NA
95	2016/09/10	NA	NA	NA	NA
96	2016/10/10	NA	NA	NA	NA
97	2016/11/10	NA	NA	NA	NA
113	2016/12/31	-1.3018	-0.4259	-1.0787	-1.2940
114	2017/01/10	-1.3670	-0.0413	-0.3334	1.0568
115	2017/01/20	-1.0886	0.9139	1.2480	0.8402
116	2017/01/31	-0.2211	0.6054	0.2192	0.3299
117	2017/02/10	0.0691	0.5113	-0.1669	1.0104
118	2017/02/20	-0.2507	1.2585	1.1370	1.5671
119	2017/02/28	0.0572	2.2190	1.6950	1.5980
120	2017/03/10	1.0225	2.0387	1.8566	0.3454
121	2017/03/20	1.5288	1.2859	0.4872	0.3763
122	2017/03/31	1.2149	0.3447	-0.2940	-0.6754
123	2017/04/10	0.6287	-0.2454	-0.7373	-0.3042
124	2017/04/20	0.1490	-0.5503	-0.9192	-0.5826
125	2017/04/30	-0.1500	-0.6513	-1.0742	-0.8764
126	2017/05/10	-0.4698	-0.9134	-1.0236	-1.0001
127	2017/06/10	NA	NA	NA	NA
128	2017/07/10	NA	NA	NA	NA
129	2017/08/10	NA	NA	NA	NA

130	2017/09/10	NA	NA	NA	NA
131	2017/10/10	NA	NA	NA	NA
132	2017/11/10	NA	NA	NA	NA
133	2017/12/10	NA	NA	NA	NA
151	2018/01/20	-1.3433	-0.8777	-0.9386	-0.6444
152	2018/01/31	-1.4173	0.2838	0.0881	0.7011
153	2018/02/10	-1.1005	1.4895	1.2383	2.2940
154	2018/02/20	-0.2478	2.0088	1.8716	2.7580
155	2018/02/28	0.7589	2.7898	1.7675	2.3713
156	2018/03/10	1.4843	1.9113	1.7358	1.4589
157	2018/03/20	1.8189	1.3958	0.9555	0.9021
158	2018/03/31	1.6235	1.0183	-0.0041	0.3454
159	2018/04/10	1.1350	0.4771	0.0934	0.4381
160	2018/04/20	0.6435	0.1737	0.5656	0.0206
161	2018/04/30	0.6523	0.4990	0.2238	-0.2114
162	2018/05/10	0.8122	-0.0826	-0.3745	-0.7372
163	2018/05/20	0.5931	-0.8159	-0.5542	-1.0620
164	2018/05/31	0.1075	-0.9458	-0.6178	-1.1239
165	2018/06/10	-0.1856	-1.0431	-0.7468	-1.2321
166	2018/06/20	-0.4728	-1.1148	-0.8884	-1.3094
167	2018/07/20	NA	NA	NA	NA
168	2018/08/20	NA	NA	NA	NA
169	2018/09/20	NA	NA	NA	NA
170	2018/10/20	NA	NA	NA	NA
171	2018/11/20	NA	NA	NA	NA
172	2018/12/20	NA	NA	NA	NA
187	2019/01/20	-1.3048	-1.0248	-1.2013	-0.9692
188	2019/01/31	-1.3018	-0.8269	-1.1900	-1.0156
189	2019/02/10	-1.3255	0.0791	-0.8259	1.0104
190	2019/02/20	-1.2604	1.4724	0.7711	1.0413
191	2019/02/28	-0.3662	1.6945	0.9864	1.1032
192	2019/03/10	0.6435	0.8501	0.0574	0.9640
193	2019/03/20	0.7471	0.8997	0.4425	0.7011
194	2019/03/31	0.4717	0.8660	0.3461	0.0051
195	2019/04/10	0.6642	0.2434	-0.2127	-0.3042
196	2019/04/20	0.5902	-0.0536	-0.7104	0.0051
197	2019/04/30	0.1638	-0.4410	-0.4998	-0.5207
198	2019/05/10	-0.1500	-0.3649	-0.4975	-0.4588
199	2019/05/20	-0.0819	-0.4627	-0.6109	-0.7372
200	2019/05/31	-0.1116	-0.9711	-0.9589	-1.3249
201	2019/06/10	NA	NA	NA	NA
202	2019/07/10	NA	NA	NA	NA
203	2019/08/10	NA	NA	NA	NA
204	2019/09/10	NA	NA	NA	NA
205	2019/10/10	NA	NA	NA	NA
206	2019/11/10	NA	NA	NA	NA

221	2019/12/31	-1.3462	-0.7323	-1.2023	-1.3558
222	2020/01/10	-1.4054	0.0725	-1.1574	0.4381
223	2020/01/20	-1.1745	1.0716	0.9194	1.1341
224	2020/01/31	-0.2803	2.1026	1.3734	0.7629
225	2020/02/10	0.5280	1.6544	1.3388	1.5362
226	2020/02/20	0.9632	2.5293	2.6252	2.7270
227	2020/02/29	1.3452	2.6278	2.0921	2.4332
228	2020/03/10	1.9048	2.1856	0.8654	2.0620
229	2020/03/20	1.8989	1.7883	1.1063	1.3351
230	2020/03/31	1.5880	1.3622	0.5610	1.3351
231	2020/04/10	1.3126	0.2649	-0.3370	-0.1805
232	2020/04/20	0.9484	-0.3309	-0.6199	-0.7372
233	2020/04/30	0.4155	-0.7512	-0.6718	-0.4279
234	2020/05/10	-0.1352	-0.6199	-0.6500	-0.9073
235	2020/05/20	-0.3899	-0.8696	-0.8055	-1.3249
236	2020/06/10	NA	NA	NA	NA
237	2020/07/10	NA	NA	NA	NA
238	2020/08/10	NA	NA	NA	NA
239	2020/09/10	NA	NA	NA	NA
240	2020/10/10	NA	NA	NA	NA
241	2020/11/10	NA	NA	NA	NA
257	2020/12/31	-1.1982	-1.1957	-1.0875	-1.3094
258	2021/01/10	-1.2604	-1.2312	-0.6903	0.4072
259	2021/01/20	-1.2722	-1.2582	1.3103	1.1650
260	2021/01/31	-1.3314	-1.3794	1.4944	2.4641
261	2021/02/10	-1.2900	-1.2987	2.1219	1.1959
262	2021/02/20	-1.2959	-1.3141	0.7958	1.4589
263	2021/02/28	-1.2189	-1.2070	-0.3334	0.6856
264	2021/03/10	-0.6919	-0.1874	-0.8645	0.0979
265	2021/03/20	0.0306	0.8425	-1.0872	0.0360
266	2021/03/31	0.6346	1.5549	-0.5545	-0.0877
267	2021/04/10	0.5931	1.4308	-0.8560	-0.7991
268	2021/04/20	0.0661	0.8608	-1.1331	-0.9847

8.6 Appendix F - Python Script to Retrieve and Spatially Subset Data

```
#!/usr/bin/python3
# -*- coding: utf-8 -*-
__author__    = "Tim Jacobs"
__copyright__ = "Copyright (c) 2020 VITO"
__license__   = "VITO NV"
__module__    = "aoi_extract_gtiff.py"
__version__   = "1.01"
__revision__ = filter(str.isdigit, "$Revision: 47442 $")
__date__     = filter(str.isalnum, "$Date: 2021-11-08 10:33:47 +0100 $")

# ---- python includes ----
import os
import re
import sys
import io
from netCDF4 import Dataset
import numpy
import shutil
import traceback
from osgeo import gdal, osr
import string
import unicodedata
import time
import subprocess

from datetime import datetime
from collections import OrderedDict

#execute-------------------------------------------------------------
# Executes a command and returns the status and output.
##@brief Function to execute a command
# @param cmd        (string)      Path to executable
# @param debug      (boolean)     Whether to log debug messages
# @param output     (boolean)     Whether to log output
# @param env        (dictionary)  The alternate env to use (allows adding dirs to existing
PATH / LD_LIBRARYPATH)
# @param shell      (boolean)     Execute command through the shell
#
# @return stat,mess (tuple)       Error code and output message
def execute_OLD(cmd,debug=None,output=True,env=None,shell=False):

    bufsize = -1 ; # 0 | 1 | x = unbuffered | line buffered | use buffer of that size

    debug = ('debug' if debug else '')
    list2cmdline = lambda x: x if isinstance(x,basestring) else subprocess.list2cmdline(x)
    if debug:
        print('execute(' + list2cmdline(cmd) + ')')
        print('=== ' + time.strftime("%Y-%m-%d %H:%M:%S",time.localtime()) + \
                                ' - START on ' + socket.gethostname())
    try:
        process =
subprocess.Popen(cmd,bufsize=bufsize,stdout=subprocess.PIPE,stderr=subprocess.STDOUT,env=env,
shell=shell);
    except Exception as e:
        print('ERROR: ',str(e))
        return (1,str(e))
```

```
    text = ''
    #make sure we enter the loop at least once; emulate do/while loop
    doWhileCondition = True
    # Get the command output and wait for the process to be finished
    while doWhileCondition:
        lines = process.stdout.readlines()
        for line in lines:
            if line:
                text = text + line
                sys.stdout.flush()
                sys.stderr.flush()
        if doWhileCondition:
            doWhileCondition = (process.poll() and process.wait()) is None

    # Get the return status from the command
    stat = process.wait()

    if text[-1:] == '\n': text = text[:-1]

    if debug:
        print('=== ' + time.strftime("%Y-%m-%d %H:%M:%S",time.localtime()) + \
                                ' - END (exit = ' + str(stat) + ')')

    return stat,text

def execute(cmd, debug=None, env=None,shell=False):
    if debug:
        print('execute(' + list2cmdline(cmd) + ')')
        print('=== ' + time.strftime("%Y-%m-%d %H:%M:%S",time.localtime()) + \
                                ' - START on ' + socket.gethostname())
    list2cmdline = lambda x: x if isinstance(x,basestring) else subprocess.list2cmdline(x)
    subprocess.check_call(cmd, stdout=sys.stdout, stderr=subprocess.STDOUT, env=env,
shell=shell)
    if debug:
        print('=== ' + time.strftime("%Y-%m-%d %H:%M:%S",time.localtime()) + \
                                ' - END (exit = ' + str(stat) + ')')

def check_status(status, message, warnStats=None):
    if status and message:
        if warnStats and status in warnStats:
            print('WARNING: ',message)
        else:
            print('ERROR: ',message)
            raise Exception(message)

def get_bands(file):
    ds = Dataset(file)
    bands = [k for k in ds.variables if len(ds.variables[k].dimensions) >= 2 and k not in
ds.dimensions]
    ds.close()
    return bands

def build_netcdf_path(file, band):
    return "NETCDF:\"" + file + "\":" + band

#FIXME: use strip_accents_unicode instead of str()?
def get_ncbandattrs(file, band):
    ds = Dataset(file)
    result = {}
```

```
    for attr in ds.variables[band].ncattrs():
        result[attr] = str(ds.variables[band].getncattr(attr))
    ds.close()
    return result

def get_ncfileattrs(nc_file):
    with Dataset(nc_file) as ds:
         return OrderedDict(sorted(ds.__dict__.items(), key=lambda t: t[0]))

# Replaces accented characters (e.g. ê) with closest match without accent (e)
# and returns the string as ascii byte string (str) instead of unicode
# See https://www.programcreek.com/python/example/4364/unicodedata.combining - Example 6
def strip_accents_unicode(input_str):
    normalized = unicodedata.normalize('NFKD', input_str)
    if normalized == input_str:
        return input_str
    else:
        return ''.join([c for c in normalized if not unicodedata.combining(c)])

#def build_tiff_filename(nc_file, band, single):
#    if single:
#        # ignore bandname
#        return os.path.splitext(os.path.basename(nc_file))[0] + ".tiff"
#    else:
#        occurence = 3 # check for third occurence
#        parts = os.path.splitext(os.path.basename(nc_file))[0].split('_')
#        return '_'.join(parts[:occurence]) + '-' + band + '_' + '_'.join(parts[occurence:])
+ ".tiff"

def str2filename(inputString):
    s=inputString
    for r in ((('I.',''),('Is.',''),("Cote d'Ivory","Ivory Coast"),('& ',''),(' ','-')):
        s=s.replace(*r)
    valid_chars = "-%s%s" % (string.ascii_letters, string.digits)
    return ''.join(c if c in valid_chars else '' for c in s).strip('-').strip()

def fix_metadata(file2fix, nc_file, bandname, productIdentifier=None, region_from=None,
region_to=None):
    ds=gdal.Open(file2fix, gdal.GA_Update)

    # get original metadata for band
    band_attrs = get_ncbandattrs(nc_file, bandname)

    # update band metadata
    band = ds.GetRasterBand(1)
    band.SetMetadata(band_attrs)

    #if("scale_factor" in band_attrs):
    #   band.SetScale(float(band_attrs["scale_factor"]))
    #if ("add_offset" in band_attrs):
    #    band.SetOffset(float(band_attrs["add_offset"]))

    if ("missing_value" in band_attrs):
        band.SetNoDataValue(float(band_attrs["missing_value"]))
    else:
        band.SetNoDataValue(float(band_attrs["_FillValue"]))

    # update identifier metadata
    #file_md = ds.GetMetadata()
```

```
    file_md = get_ncfileattrs(nc_file)

    for key in file_md:
        if '#' in key:
            if key.startswith('NC_GLOBAL'):
                new_key=key[key.rfind('#')+1:]
                file_md[new_key]=file_md[key]
                file_md.pop(key,None)
            else:
                # lat#, lon#, crs#, time# keys
                file_md.pop(key,None)
        if key.startswith('NETCDF_'):
            file_md.pop(key,None)

    if 'parent_identifier' not in file_md:

file_md['parent_identifier']=file_md['identifier'][0:file_md['identifier'].rfind(':')]

    if productIdentifier:
        file_md['identifier']=file_md['parent_identifier']+':'+productIdentifier

    if region_from and region_to:
        #file_md['identifier'] = file_md['parent_identifier'] +
file_md['identifier'][file_md['identifier'].rfind(':'):].replace(region_from,region_to)
        file_md['title'] = file_md['title'].replace(region_from,region_to)
    else:
        file_md['identifier'] = file_md['parent_identifier'] +
file_md['identifier'][file_md['identifier'].rfind(':'):]

    if region_to:
        file_md['region_name'] = region_to

    # force https
    if 'references' in file_md:
        file_md['references'] = file_md['references'].replace('http','https')
    else:
        prodtype=os.path.basename(nc_file).split('_')[2].split('-')[0]
        re_match = re.search("\d", prodtype)
        #if no digit in prodtype, use full string in lowercase
        var = prodtype.lower()
        if re_match:
            #if digit in prodtype string, then we use all characters up to first digit
            var=var[0:re_match.start()]
        file_md['references']='https://land.copernicus.eu/global/products/' + var
        #print('REFERENCES: '+file_md['references'])

    # overwrite history
    file_md['history'] = datetime.now().strftime("%Y-%m-%d") + ' -- Custom order'

    # overwrite copyright
    file_md['copyright'] = 'Copernicus Service information '+ str(datetime.now().year)

    # Python3 does not have separate str/unicode types, only raw bytes or str
    # JIRA PVF-1401
    for k in file_md:
        if not isinstance(file_md[k],str):
            #print('POP KEY',k)
            file_md.pop(k,None)
```

```
    # JIRA PVF-1428: gdal's SetMetadata doesn't allow unicode strings,
    # only byte strings (str), in dictionary. So we try to convert.
    # For LST products, eg 'Português' needs to be replaced with 'Portugues'
    # For SCE 1km products, a EN DASH (\u2013, looks like long hyphen) is to be replaced with
a regular one (-)

    #FIXME
    # JIRA PVF-1441
    # For SWE products, the netCDF4.Dataset.getncattr() returns \ufffd (unknown char)
    # when decoding the 'comment' attribute, while h5py returns the degree symbol (°)
correctly.
    # For time being, we'll just remove any \ufffd's
    # Ideally, get_ncfileattrs should be adapted to fall back to h5py, so that ° can be
retained or
    # replaced with eg 'degree' in full. Or a regexp could look for [digits]°[NESW] pattern.

    for x in file_md:
        if not isinstance(file_md[x],str):
            file_md[x]=file_md[x].replace('\u2013','-')
            file_md[x]=file_md[x].replace('\ufffd','')
            #print('key',x,'value',file_md[x])
            try:
                file_md[x]=str(file_md[x])
            except UnicodeEncodeError:
                #Some special chars cannot be encoded as ascii string
                file_md[x]=strip_accents_unicode(file_md[x])
                #print('Fixed codec for metadata key',x,'value',str(file_md[x]))

    #print(file_md)
    #del file_md['comment']
    ds.SetMetadata(file_md)

    proj=ds.GetProjection()
    if not proj or len(proj)==0:
        sr=osr.SpatialReference()
        sr.ImportFromEPSG(4326)
        ds.SetProjection(sr.ExportToWkt())

    #Copy CRS and GeoTransform from input netCDF subdataset
    #gd=gdal.Open(build_netcdf_path(nc_file,bandname),gdal.GA_ReadOnly)
    #ds.SetProjection(gd.GetProjection())
    #ds.SetGeoTransform(gd.GetGeoTransform())
    #gd = None

    band = None
    ds = None

def
extract_aoi_from_netcdf(inputFile,outputDir,bandsSelected=None,aoi_proj_coordinates=None,aoi_
name='CUSTOM'):
    # /some/path/to/c_gls_...Vx.y.z.n
    nc_file            = inputFile
    # c_gls_...Vx.y.z.nc
    nc_file_base       = os.path.basename(nc_file)
    # c_gls_...Vx.y.z
    nc_file_base_noExt = os.path.splitext(nc_file_base)[0]

    print("Processing: " + nc_file_base)
```

```
    bands = get_bands(nc_file)

    if bandsSelected:
        bands = set(bands).intersection(set(bandsSelected))

    print('Bands to extract: ' + ', '.join(bands))

    # Split input filename into parts
    filename_elements = nc_file_base_noExt.split('_')

    product_roi = filename_elements[4]

    # Replace ROI name in filename for any output files
    filename_elements[4] = aoi_name

    # Retain the variable name from filename, so we can keep appending layer (band) names
    variable = filename_elements[2]

    # Retain product version
    product_version = filename_elements[6]
    if len(product_version.split('.'))==2:
        product_version=product_version+'.1'
        filename_elements[6]=product_version

    for band in bands:
        print('-- Extracting band',band,'from product',nc_file_base)
        filename_elements[2] = variable + '-' + band.replace('_','-')
        outfile_base_noExt = '_'.join(filename_elements)
        productID = outfile_base_noExt.replace('c_gls_','')

        outputVrtBaseName = outfile_base_noExt + '.vrt'
        outputVrtFileName = os.path.join(outputDir,outputVrtBaseName)

        outputfile = os.path.join(outputDir, outfile_base_noExt+'.img')

        cmd = ['gdal_translate']
        #cmd += ['-of', 'GTIFF' ]
        cmd += ['-of', 'ENVI' ]
       # requires GDAL v2.2 or higher
        cmd += ['-oo', 'HONOUR_VALID_RANGE=NO']

        #cmd += ['-co','COMPRESS=LZW']
        #cmd += ['-co','TILED=YES']
        #cmd += ['-co','BIGTIFF=IF_NEEDED']

        # LIE 250m Baltic is incorrectly written
        #if variable.lower().startswith('lie') and product_roi.lower() == 'baltic' and
product_version.lower() == 'v1.0.1':
        #    cmd += ['--config','GDAL_NETCDF_BOTTOMUP','NO']

        if aoi_proj_coordinates:
            west  = float(aoi_proj_coordinates[0])
            north = float(aoi_proj_coordinates[1])
            east  = float(aoi_proj_coordinates[2])
            south = float(aoi_proj_coordinates[3])
            cmd += ['-projwin'] + [str(west),str(north),str(east),str(south)]

        cmd += [build_netcdf_path(nc_file, band), outputfile]
```

```
        print('Executing: ', ' '.join(cmd))

        execute(cmd)
        #status,message = execute(cmd)
        #check_status(status,message)

        print('Updating Metadata')
        #ds = gdal.Open(outputfile, gdal.GA_Update)
        fix_metadata(outputfile, nc_file, band, productIdentifier=productID,
region_from=product_roi, region_to=aoi_name)
        #ds = None

    print('All Done!')

if __name__ == '__main__':

    import argparse
    myparser=argparse.ArgumentParser(description='Extract AOIs from CGLS products in netCDF
format and store them in GeoTIFF.')

    def csv_list_parser(string):
        return string.split(',')

    def dir_write_check(string):
        cond = os.path.isdir(string) and \
            os.access(string,os.W_OK)
        if cond:
            return string
        else:
            print('Directory ',string,' must be an existing, writable directory for storing
the GeoTIFF outputs.')
            myparser.error()

    # argparse FileType leaves the file open,
    # it is preferred to just check existance and readability
    # and we add some filename checks
    def file_read_check(string):
        cond = os.path.isfile(string) and \
            os.access(string,os.R_OK) and \
            os.path.basename(string).lower().startswith('c_gls_') and \
            string.lower().endswith('.nc')
        if cond:
            return string
        else:
            print('File ',string,' must be an existing, readable netCDF input file.')
            myparser.error()

    #output directory
    myparser.add_argument('outputDir',metavar='output_directory',help='output directory for
extracted GeoTIFF', nargs=1,type=dir_write_check)

    #input file
    myparser.add_argument('inputFile',metavar='input_file',help='input netCDF
file',nargs='+',type=file_read_check)

    # optional list of bands
    myparser.add_argument('--bands','-b',metavar='band1 band2 ... band_n',help='list of
netCDF bands',nargs='+',default=None, dest='bands')
```

```
    # option aoi coordinates with or without name
    myparser.add_argument('--aoi','--projwin',metavar='float', help='AOI coordinates: ULx ULy
LRx LRy',nargs=4,type=float )
    myparser.add_argument('--aoi_name','-n',metavar='name',help='name for the custom
AOI',nargs=1, dest='aoi_name', default='CUSTOM')

    #vars() turns the argparse.Namespace into a dictionary
    myargs=vars(myparser.parse_args())

    # for debugging argparser
    #for k in myargs:
    #    print('arg',k,'=',myargs[k])
    #exit(0)

    for infile in myargs['inputFile']:

extract_aoi_from_netcdf(infile,myargs['outputDir'][0],bandsSelected=myargs['bands'],aoi_proj_
coordinates=myargs['aoi'],aoi_name=myargs['aoi_name'][0])
```

PYTHON SCRIPT REFERENCE: (Jacobs and VITO NV, 2020)

8.7 Appendix G - Linux Bash Scripts

8.7.1 *f*APAR Data Access Script

```
#!/bin/bash
cd /data/MTDA/BIOPAR/BioPar_FAPAR300_V1_Global/
#find . -type f ! -name '.nc' -print -exec sh -c 'cdo info {} > {}.nc' \;
x=( $(find . -name "*.nc"));
#echo "${x[@]}" # Print the names of the elements in x
for item in "${x[@]}";
#{ echo "$item"; }
do
        #!<python3>
        # Karoo 1 - fAPAR ENVI
        python3 /home/jasonj9895/Desktop/Biomass_Working_Folder/1.aoi_extract_envi.py --bands
FAPAR --aoi 25.02530 -31.4183138 25.03342 -31.4264592 --aoi_name Karoo-1-envi
/home/jasonj9895/Desktop/Biomass_Working_Folder/fAPAR_Data/Karoo_1_ENVI "$item"
done
echo ${#x[@]} # No. of elements in x
```

8.7.2 *f*APAR Update Statistics Script

```
#!/bin/bash
# Update Statistics on all images

cd /home/jasonj9895/Desktop/Biomass_Working_Folder/fAPAR_Data/Karoo_1_ENVI
for j in $(ls *.img); do gdal_edit.py -stats -a_nodata 9999 $j ;done
```

8.7.3 *f*APAR RT2 & RT6 Data Extraction Script

```
# Extract data for RT2 & RT6 Datasets
######################### RT2 & 6 ######################################

cd /home/jasonj9895/Desktop/Biomass_Working_Folder/fAPAR_Data/Karoo_1_ENVI/RT_2_6

find -name \*.img -printf "%p\n" >
/home/jasonj9895/Desktop/Biomass_Working_Folder/fAPAR_Data/Karoo_1_ENVI/RT_2_6/k1_fAPAR_envi_
filenames_RT_2_6.txt

for j in $(ls *.img); do gdalinfo $j
>>/home/jasonj9895/Desktop/Biomass_Working_Folder/fAPAR_Data/Karoo_1_ENVI/RT_2_6/k1_fAPAR_env
i_info_RT_2_6.txt ;done

for j in $(ls *.img); do gdalinfo $j  |grep time_coverage_end
>>/home/jasonj9895/Desktop/Biomass_Working_Folder/fAPAR_Data/Karoo_1_ENVI/RT_2_6/k1_fAPAR_env
i_time_RT_2_6.txt ;done

for j in $(ls *.img); do gdalinfo $j  |grep STATISTICS_MEAN
>>/home/jasonj9895/Desktop/Biomass_Working_Folder/fAPAR_Data/Karoo_1_ENVI/RT_2_6/k1_fAPAR_env
i_mean_RT_2_6.txt ;done

for j in $(ls *.img); do gdalinfo $j  |grep STATISTICS_STDDEV
>>/home/jasonj9895/Desktop/Biomass_Working_Folder/fAPAR_Data/Karoo_1_ENVI/RT_2_6/k1_fAPAR_env
i_std_RT_2_6.txt ;done
```

8.7.4 GDMP Data Access Script

```
#### GDMP ####
#!/bin/bash
cd /data/MTDA/BIOPAR/BioPar_GDMP300_V1_Global/
x=( $(find . -name "*.nc") );
#echo "${x[@]}" # Print the names of the elements in x
#echo ${#x[@]} # No. of elements in x
for item in "${x[@]}";
do
        #!<python3>
        # Karoo 1 - GDMP
        python3 /home/jasonj9895/Desktop/Biomass_Working_Folder/5.aoi_extract_envi.py --bands
GDMP --aoi 25.02530 -31.4183138 25.03342 -31.4264592 --aoi_name Karoo-1_envi
/home/jasonj9895/Desktop/Biomass_Working_Folder/GDMP_Data/Karoo_1_ENVI "$item"
        # Karoo 2 - GDMP
        #python3 /home/jasonj9895/Desktop/Biomass_Working_Folder/5.aoi_extract_gtiff.py --
bands GDMP --aoi 25.0112074 -31.427081 25.0207804 -31.43474 --aoi_name Karoo-2
/home/jasonj9895/Desktop/Biomass_Working_Folder/GDMP_Data/Karoo_2 "$item"
done

echo ${#x[@]} # No. of elements in x
```

8.7.5 GDMP Update Statistics Script

```
#!/bin/bash
# Update Statistics on all images

cd /home/jasonj9895/Desktop/Biomass_Working_Folder/GDMP_Data/Karoo_1_ENVI
for j in $(ls *.img); do gdal_edit.py -stats -a_nodata 9999 $j ;done
```

8.7.6 GDMP RT5 & RT6 Data Extraction Script

```
# Extract data for RT5 & RT6 Datasets
################## RT 5 & 6 ######################

cd /home/jasonj9895/Desktop/Biomass_Working_Folder/GDMP_Data/Karoo_1_ENVI/RT_5_6
find -name \*.img >
/home/jasonj9895/Desktop/Biomass_Working_Folder/GDMP_Data/Karoo_1_ENVI/RT_5_6/k1_GDMP_envi_fi
lenames_RT_5_6.txt

for j in $(ls *.img); do gdalinfo $j
>>/home/jasonj9895/Desktop/Biomass_Working_Folder/GDMP_Data/Karoo_1_ENVI/RT_5_6/k1_GDMP_envi_
info_RT_5_6.txt ;done

for j in $(ls *.img); do gdalinfo $j  |grep time_coverage_end
>>/home/jasonj9895/Desktop/Biomass_Working_Folder/GDMP_Data/Karoo_1_ENVI/RT_5_6/k1_GDMP_envi_
time_RT_5_6.txt ;done

for j in $(ls *.img); do gdalinfo $j  |grep STATISTICS_MEAN
>>/home/jasonj9895/Desktop/Biomass_Working_Folder/GDMP_Data/Karoo_1_ENVI/RT_5_6/k1_GDMP_envi_
mean_RT_5_6.txt ;done

for j in $(ls *.img); do gdalinfo $j  |grep STATISTICS_STDDEV
>>/home/jasonj9895/Desktop/Biomass_Working_Folder/GDMP_Data/Karoo_1_ENVI/RT_5_6/k1_GDMP_envi_
std_RT_5_6.txt ;done
```

8.8 Appendix H - R Analysis Scripts

8.8.1 Full (All Phenophases) Dataset Script

```
####### FULL DATASET ######

###Based on
#### Script to plot MODIS collection 6 subset data
#### Alison Boyer - ORNL DAAC
#### April 2017
#### https://github.com/ornldaac/modis/blob/master/modis-global-fixed-statistics.ipynb

#### Load required packages. You may need to run install.packages('package') if you have not
used these before.
library(ggplot2)
library(viridisLite)
library(readr)
library(dplyr)
#library(bigleaf) # To check xcel conversion of umol/m2/sec to gC/m2/day
#umolCO2.to.gC(0.0365)  # gC m-2 d-1
## gC/m²/day = (µmol/m²/s * 12.01 g/mol) * (24 * 60 * 60 seconds/day) / 1,000,000 / Area

#### Read in data
getwd()
setwd("G:/AAA.Google_Drive/THESIS_BIOMASS_ANOMALIES/R_Processing")
All_Data <- read_csv("all_data_sub_SEPTEMBER.csv")
#summary(All_Data) # get basic summary information about the data
#View(All_Data)
# "Converted" columns are MODIS_mean from kilograms to grams and factor applied
# gC/m²/day = (µmol CO2/m²/s * 12.01 g/mol) * (24 * 60 * 60 seconds/day) / 1,000,000 / Area

str(All_Data)
num_observations <- as.data.frame(colSums(!is.na(All_Data)))
View(num_observations)

# Format dates and add to data frame
dates <- as.Date(All_Data$Date)
years <- format(dates, format = "%Y")
n_yrs <- length(unique(years)) # calculate number of years in subset
months <- format(dates, format = "%m")
monthAbbrev <- format(dates, format = "%b")

# find the 95% confidence intervals around the mean i.e -3 to +3 standard deviation - include
outliers
y1 <- All_Data$MODISGPP_gCm2day + (3 * All_Data$MODISGPP_SD_gCm2day)
y2 <- All_Data$MODISGPP_gCm2day - (3 * All_Data$MODISGPP_SD_gCm2day)

# add to main data frame
All_Data <- cbind(All_Data, months, years, y1, y2)

#### Plotting
names(All_Data)
#### Figure 1: mean and variation time series: subset mean is a dark line and 95% CI is a
shaded region
F1_FULL <- ggplot(All_Data, aes(x=as.Date(Date), y = MODISGPP_gCm2day)) +
  geom_ribbon(aes(ymin=y2, ymax=y1), fill = "darkgray") + # draw the shaded area for 95% CI
  geom_line(colour = "#2D708EFF", aes(y = MODISGPP_gCm2day)) + # draw the line for the mean
value
```

```
  ggtitle("MODIS Mean GPP and 95% CI") + # add a title
  ylab(expression(paste("MODIS Mean GPP 500m (gC/m"^{2} * "/day)"))) + # add a y-axis label
  xlab("Date") + # add a x-axis label
  theme_bw() + # set the plot theme
  theme(plot.title = element_text(lineheight=.8, face="bold", size=20), axis.title.x =
element_text(face="bold", size=20), axis.title.y = element_text(face="bold", size=20),
axis.text.x = element_text(size=16), axis.text.y = element_text(size=16)) # set optional
theme elements
F1_FULL

#### Figure 2: mean monthly values for the subset
F2_FULL <- ggplot(All_Data, aes(factor(months), MODISGPP_gCm2day)) +
  geom_boxplot(fill="#2D708EFF") +
  geom_jitter(width = 0) +
  ggtitle("MODIS Mean GPP by Month") +
  scale_x_discrete(breaks= months, labels=monthAbbrev) +
  xlab("Month") +
  ylab(expression(paste("MODIS Mean GPP 500m (gC/m"^{2} * "/day)"))) +
  theme_bw() +
  theme(plot.title = element_text(lineheight=.8, face="bold", size=20),
legend.position="none", axis.title.x = element_text(face="bold", size=20), axis.text.x =
element_text(angle=90, vjust=0.5, size=16), axis.title.y = element_text(face="bold",
size=20), axis.text.y = element_text(size=16))
F2_FULL

#### Figure 3: mean yearly values for the subset
F3_FULL <- ggplot(All_Data, aes(factor(years), MODISGPP_gCm2day)) +
  geom_boxplot(fill="#2D708EFF") +
  geom_jitter(width = 0) +
  ggtitle("MODIS Mean GPP by Year") +
  scale_x_discrete(breaks= years, labels=years) +
  xlab("Year") +
  ylab(expression(paste("MODIS Mean GPP 500m (gC/m"^{2} * "/day)"))) +
  theme_bw() +
  theme(plot.title = element_text(lineheight=.8, face="bold", size=20),
legend.position="none", axis.title.x = element_text(face="bold", size=20), axis.text.x =
element_text(angle=90, vjust=0.5, size=16), axis.title.y = element_text(face="bold",
size=20), axis.text.y = element_text(size=16))
F3_FULL
##############################
mydata_FULL <- All_Data

#### Interim Dataset Statistics
summary(mydata_FULL$fAPAR*(1/250))
sd(mydata_FULL$fAPAR*(1/250), na.rm=TRUE)
print(sum(!is.na(mydata_FULL$fAPAR)))

summary(mydata_FULL$GDMP)
sd(mydata_FULL$GDMP, na.rm=TRUE)
print(sum(!is.na(mydata_FULL$GDMP)))

summary(mydata_FULL$mean_GPP_DT_gCm2day)
sd(mydata_FULL$mean_GPP_DT_gCm2day, na.rm=TRUE)
print(sum(!is.na(mydata_FULL$mean_GPP_DT_gCm2day)))

summary(mydata_FULL$MODISGPP_gCm2day)
sd(mydata_FULL$MODISGPP_gCm2day, na.rm=TRUE)
print(sum(!is.na(mydata_FULL$MODISGPP_gCm2day)))
```

```
##### END FOR FULL DATASETS - AUGUST 2023 ####

# fAPAR to GDMP #
cor.test(mydata_FULL$fAPAR*(1/250), mydata_FULL$GDMP, use = "complete.obs")
non_missing_indices <- complete.cases(mydata_FULL$fAPAR, mydata_FULL$GDMP)
n <- sum(non_missing_indices)
print(n)

cor.test(mydata_FULL$fAPAR*(1/250), mydata_FULL$mean_GPP_DT_gCm2day, use = "complete.obs")
non_missing_indices <- complete.cases(mydata_FULL$fAPAR, mydata_FULL$mean_GPP_DT_gCm2day)
n <- sum(non_missing_indices)
print(n)

cor.test(mydata_FULL$GDMP, mydata_FULL$mean_GPP_DT_gCm2day, use = "complete.obs")
non_missing_indices <- complete.cases(mydata_FULL$GDMP, mydata_FULL$mean_GPP_DT_gCm2day)
n <- sum(non_missing_indices)
print(n)

cor.test(mydata_FULL$MODISGPP_gCm2day, mydata_FULL$fAPAR*(1/250), use = "complete.obs")
non_missing_indices <- complete.cases(mydata_FULL$MODISGPP_gCm2day,
mydata_FULL$fAPAR*(1/250))
n <- sum(non_missing_indices)
print(n)

cor.test(mydata_FULL$MODISGPP_gCm2day, mydata_FULL$GDMP, use = "complete.obs")
non_missing_indices <- complete.cases(mydata_FULL$MODISGPP_gCm2day, mydata_FULL$GDMP)
n <- sum(non_missing_indices)
print(n)

cor.test(mydata_FULL$MODISGPP_gCm2day, mydata_FULL$mean_GPP_DT_gCm2day, use = "complete.obs")
non_missing_indices <- complete.cases(mydata_FULL$MODISGPP_gCm2day,
mydata_FULL$mean_GPP_DT_gCm2day)
n <- sum(non_missing_indices)
print(n)

save.image("G:/AAA.Google_Drive/THESIS_BIOMASS_ANOMALIES/R_Processing/Pheno_sub_data_sub_SEPT
EMBER.ANALYSIS.RData")

### END OF ALL DATA ANAYLSES ###

### Citations ###
citation()
version$version.string
citation("ggplot2")
citation("dplyr")
citation("pROC")
citation('moments')
citation("jtools")
```

8.8.2 Phenophase Subset (Greenup – Dormancy) Analysis Dataset Script

```
####### PHENOPHASE SUBSET DATASET ######

###Based on Script to plot MODIS collection 6 subset data
#### Alison Boyer - ORNL DAAC - April 2017
#### https://github.com/ornldaac/modis/blob/master/modis-global-fixed-statistics.ipynb

#### Load required packages
library(ggplot2)
library(viridisLite)
library(readr)
library(dplyr)
# library(bigleaf) - to check xcel conversion of umol/m2/sec to gC/m2/day
# umolCO2.to.gC(0.0365)  # gC m-2 d-1
# gC/m²/day = (µmol CO2/m²/s * 12.01 g/mol) * (24 * 60 * 60 seconds/day) / 1,000,000 / Area

#### Read in data
getwd()
#setwd("G:/AAA.Google_Drive/THESIS_BIOMASS_ANOMALIES/R_Processing")
# For REANALYSIS, copy placed in wd.
subset_pheno <- read_csv("all_data_sub_GREENUP_to_DORMANCY_SEPTEMBER.csv")
#setwd("G:/AAA.AAA.MASTER_LUND_MSC/Semester_5/THESIS BIOMASS
ANOMALIES/THESIS/AAA.AAA.REVISED/REANALYSIS")
str(subset_pheno)
num_observations <- as.data.frame(colSums(!is.na(subset_pheno)))
View(num_observations)

# Format dates and add to data frame
dates_pheno <- as.Date(subset_pheno$Date)
years_pheno <- format(dates_pheno, format = "%Y")
n_yrs_pheno <- length(unique(years_pheno)) # calculate number of years in subset
months_pheno <- format(dates_pheno, format = "%m")
monthAbbrev_pheno <- format(dates_pheno, format = "%b")

names(subset_pheno)

# find the 95% confidence intervals around the mean i.e -3 to +3 standard deviation - include
outliers
y1_pheno <- subset_pheno$MODISGPP_gCm2day + (3 * subset_pheno$MODISGPP_SD_gCm2day)
y2_pheno <- subset_pheno$MODISGPP_gCm2day - (3 * subset_pheno$MODISGPP_SD_gCm2day)

# add to main data frame
subset_pheno <- cbind(subset_pheno, months_pheno, years_pheno, y1_pheno, y2_pheno)

### Analysis ###
# Calculate fAPAR from fAPAR_DN - apply factor
subset_pheno$fAPAR_unitless <- subset_pheno$fAPAR * 0.004

## Descriptive Statistics ###
summary(subset_pheno$fAPAR_unitless, na.rm = TRUE)
sd(subset_pheno$fAPAR_unitless, na.rm=TRUE)
print(sum(!is.na(subset_pheno$fAPAR_unitless)))

summary(subset_pheno$GDMP, na.rm=TRUE)
sd(subset_pheno$GDMP, na.rm=TRUE)
print(sum(!is.na(subset_pheno$GDMP)))
```

```
summary(subset_pheno$mean_GPP_DT_gCm2day)
sd(subset_pheno$mean_GPP_DT_gCm2day, na.rm=TRUE)
print(sum(!is.na(subset_pheno$mean_GPP_DT_gCm2day)))

summary(subset_pheno$MODISGPP_gCm2day)
sd(subset_pheno$MODISGPP_gCm2day, na.rm=TRUE)
print(sum(!is.na(subset_pheno$MODISGPP_gCm2day)))

# Calculate z-scores
fAPAR_mean <- mean(subset_pheno$fAPAR_unitless, na.rm=TRUE)
fAPAR_sd <- sd(subset_pheno$fAPAR_unitless, na.rm=TRUE)
fAPAR_z_scores <- ((subset_pheno$fAPAR_unitless - fAPAR_mean) / (fAPAR_sd))
subset_pheno$fAPAR_z <- fAPAR_z_scores

GDMP_mean <- mean(subset_pheno$GDMP, na.rm=TRUE)
GDMP_sd <- sd(subset_pheno$GDMP, na.rm=TRUE)
GDMP_z_scores <- ((subset_pheno$GDMP - GDMP_mean) / (GDMP_sd))
subset_pheno$GDMP_z <- GDMP_z_scores

GPP_EC_mean <- mean(subset_pheno$mean_GPP_DT_gCm2day, na.rm=TRUE)
GPP_EC_sd <- sd(subset_pheno$mean_GPP_DT_gCm2day, na.rm=TRUE)
GPP_EC_z_scores <- ((subset_pheno$mean_GPP_DT_gCm2day - GPP_EC_mean) / (GPP_EC_sd))
subset_pheno$GPP_EC_z <- GPP_EC_z_scores

MODIS_GPP_mean <- mean(subset_pheno$MODISGPP_gCm2day, na.rm=TRUE)
MODIS_GPP_sd <- sd(subset_pheno$MODISGPP_gCm2day, na.rm=TRUE)
MODIS_GPP_z_scores <- ((subset_pheno$MODISGPP_gCm2day - MODIS_GPP_mean) / (MODIS_GPP_sd))
subset_pheno$MODIS_GPP_z <- MODIS_GPP_z_scores

# Export to csv for z-scores data
file_path <- "subset_NOVEMBER_Z_SCORES_export.csv"
write.csv(subset_pheno, file = file_path, row.names = FALSE)

#### Anomaly (z-scores) Dataset Statistics
summary(subset_pheno$fAPAR_z)
sd(subset_pheno$fAPAR_z, na.rm=TRUE)
print(sum(!is.na(subset_pheno$fAPAR_z)))

summary(subset_pheno$GDMP_z)
sd(subset_pheno$GDMP_z, na.rm=TRUE)
print(sum(!is.na(subset_pheno$GDMP_z)))

summary(subset_pheno$GPP_EC_z)
sd(subset_pheno$GPP_EC_z, na.rm=TRUE)
print(sum(!is.na(subset_pheno$GPP_EC_z)))

summary(subset_pheno$MODIS_GPP_z)
sd(subset_pheno$MODIS_GPP_z, na.rm=TRUE)
print(sum(!is.na(subset_pheno$MODIS_GPP_z)))

### Correlation Tests ###
# fAPAR to GDMP #
cor.test(subset_pheno$fAPAR_z, subset_pheno$GDMP_z, use = "complete.obs")
non_missing_indices <- complete.cases(subset_pheno$fAPAR_z, subset_pheno$GDMP_z)
n <- sum(non_missing_indices)
print(n)

# fAPAR to GPP_EC #
cor.test(subset_pheno$fAPAR_z, subset_pheno$GPP_EC_z, use = "complete.obs")
```

```
non_missing_indices <- complete.cases(subset_pheno$fAPAR_z, subset_pheno$GPP_EC_z)
n <- sum(non_missing_indices)
print(n)

# GDMP to GPP_EC #
cor.test(subset_pheno$GDMP_z, subset_pheno$GPP_EC_z, use = "complete.obs")
non_missing_indices <- complete.cases(subset_pheno$GDMP_z, subset_pheno$GPP_EC_z)
n <- sum(non_missing_indices)
print(n)

# MODIS_GPP to fAPAR
cor.test(subset_pheno$MODIS_GPP_z, subset_pheno$fAPAR_z, use = "complete.obs")
non_missing_indices <- complete.cases(subset_pheno$MODIS_GPP_z, subset_pheno$fAPAR_z)
n <- sum(non_missing_indices)
print(n)

# MODIS_GPP to GDMP
cor.test(subset_pheno$MODIS_GPP_z, subset_pheno$GDMP_z, use = "complete.obs")
non_missing_indices <- complete.cases(subset_pheno$MODIS_GPP_z, subset_pheno$GDMP_z)
n <- sum(non_missing_indices)
print(n)

# MODIS_GPP to GPP_EC
cor.test(subset_pheno$MODIS_GPP_z, subset_pheno$GPP_EC_z, use = "complete.obs")
non_missing_indices <- complete.cases(subset_pheno$MODIS_GPP_z, subset_pheno$GPP_EC_z)
n <- sum(non_missing_indices)
print(n)

#### PLOTS ####
library(gridExtra)

# Create the summary table
summary_table <- subset_pheno %>%
  group_by(months_pheno) %>%
  summarise(
    Num_Observations_fAPAR = sum(!is.na(fAPAR), na.rm=TRUE),
    Num_Observations_GDMP = sum(!is.na(GDMP), na.rm=TRUE),
    Num_Observations_mean_GPP = sum(!is.na(mean_GPP_DT_gCm2day), na.rm=TRUE),
    Num_Observations_MODIS_mean = sum(!is.na(MODISGPP_gCm2day), na.rm=TRUE)
  )
print(summary_table)

summary_table_year <- subset_pheno %>%
  group_by(years_pheno) %>%
  summarise(
    Num_Observations_fAPAR = sum(!is.na(fAPAR), na.rm=TRUE),
    Num_Observations_GDMP = sum(!is.na(GDMP), na.rm=TRUE),
    Num_Observations_mean_GPP = sum(!is.na(mean_GPP_DT_gCm2day), na.rm=TRUE),
    Num_Observations_MODIS_mean = sum(!is.na(MODISGPP_gCm2day), na.rm=TRUE)
  )
print(summary_table_year)

### Monthly Plots #####
# Plotting Monthly Values of fAPAR
# Figure 1: mean monthly values for the fAPAR subset

F1 <- ggplot(subset_pheno, aes(x = factor(months_pheno), fAPAR_unitless)) +
  geom_boxplot(fill = "#2D708EFF") +
  geom_jitter(width = 0) +
```

```
  ggtitle("fAPAR Phenophase Subset Values by Month") +
  scale_x_discrete(breaks = months_pheno, labels = monthAbbrev_pheno) +
  xlab("Month") +
  ylab("fAPAR (Unitless)") +
  theme_bw() +
  theme(
    plot.title = element_text(lineheight = 0.8, face = "bold", size = 14),
    legend.position = "none",
    axis.title.x = element_text(face = "bold", size = 12),
    axis.text.x = element_text(angle = 90, vjust = 0.5, size = 14),
    axis.title.y = element_text(face = "bold", size = 12),
    axis.text.y = element_text(size = 12)) +
  guides(fill = "none")

# Plotting Monthly Values of GDMP
# Figure 2: mean monthly values for the subset
F2 <- ggplot(subset_pheno, aes(factor(months_pheno), GDMP)) +
  geom_boxplot(fill="#2D708EFF") +
  geom_jitter(width = 0) +
  ggtitle("GDMP Phenophase Subset Values by Month") +
  scale_x_discrete(breaks= months_pheno, labels=monthAbbrev_pheno) +
  xlab("Month") +
  ylab(expression(paste("GDMP - kgDM/ha/day"))) +
  theme_bw() +
  theme(plot.title = element_text(lineheight=.8, face="bold", size=14),
        legend.position="none",
        axis.title.x = element_text(face="bold", size=12),
        axis.text.x = element_text(angle=90, vjust=0.5, size=14),
        axis.title.y = element_text(face="bold", size=12),
        axis.text.y = element_text(size=12)) +
  guides(fill="none")

# Plotting Monthly Values of GPP_EC
# Figure 3: mean monthly values for the subset
F3 <- ggplot(subset_pheno, aes(factor(months_pheno), mean_GPP_DT_gCm2day)) +
  geom_boxplot(fill="#2D708EFF") +
  geom_jitter(width = 0) +
  ggtitle("GPP_EC Phenophase Subset Values by Month") +
  scale_x_discrete(breaks= months_pheno, labels=monthAbbrev_pheno) +
  xlab("Month") +
  ylab(expression(paste("GPP EC - gC/m"^{2} * "/day"))) +
  theme_bw() +
  theme(plot.title = element_text(lineheight=.8, face="bold", size=14),
        legend.position="none",
        axis.title.x = element_text(face="bold", size=12),
        axis.text.x = element_text(angle=90, vjust=0.5, size=14),
        axis.title.y = element_text(face="bold", size=12),
        axis.text.y = element_text(size=12)) +
  guides(fill="none")

#### Plotting Monthly Values of MODIS GPP
#### Figure 4: mean monthly values for the subset
F4 <- ggplot(subset_pheno, aes(factor(months_pheno), MODISGPP_gCm2day)) +
  geom_boxplot(fill="#2D708EFF") +
  geom_jitter(width = 0) +
  ggtitle("MODIS Phenophase Subset Values by Month") +
  scale_x_discrete(breaks= months_pheno, labels=monthAbbrev_pheno) +
  xlab("Month") +
  ylab(expression(paste("MODIS GPP 500m - gC/m"^{2} * "/day"))) +
```

```
  theme_bw() +
  theme(plot.title = element_text(lineheight=.8, face="bold", size=14),
        legend.position="none", axis.title.x = element_text(face="bold", size=14),
        axis.text.x = element_text(angle=90, vjust=0.5, size=14),
        axis.title.y = element_text(face="bold", size=12),
        axis.text.y = element_text(size=12)) +
  guides(fill="none")

# Combine all plots into one graphic
combined_plot <- grid.arrange(F1, F2, F3, F4, ncol = 2)

### Yearly Plots ########
# Plotting yearly Values of fAPAR
# Figure 5: mean yearly values for the subset
F5 <- ggplot(subset_pheno, aes(factor(years_pheno), fAPAR_unitless)) +
  geom_boxplot(fill="#2D708EFF") +
  geom_jitter(width = 0) +
  ggtitle("fAPAR Phenophase Subset Values by Year") +
  scale_x_discrete(breaks= years_pheno, labels=years_pheno) +
  xlab("Year") +
  ylab(expression(paste("fAPAR (Unitless) "))) +
  theme_bw() +
  theme(plot.title = element_text(lineheight=.8, face="bold", size=14),
        legend.position="none",
        axis.title.x = element_text(face="bold", size=12),
        axis.text.x = element_text(angle=90, vjust=0.5, size=14),
        axis.title.y = element_text(face="bold", size=12),
        axis.text.y = element_text(size=12))

# Plotting yearly Values of GDMP
# Figure 6: mean yearly values for the subset
F6 <- ggplot(subset_pheno, aes(factor(years_pheno), GDMP)) +
  geom_boxplot(fill="#2D708EFF") +
  geom_jitter(width = 0) +
  ggtitle("GDMP Phenophase Subset Values by Year") +
  scale_x_discrete(breaks= years_pheno, labels=years_pheno) +
  xlab("Year") +
  ylab(expression(paste("GDMP - kgDM/ha/day"))) +
  theme_bw() +
  theme(plot.title = element_text(lineheight=.8, face="bold", size=14),
        legend.position="none",
        axis.title.x = element_text(face="bold", size=12),
        axis.text.x = element_text(angle=90, vjust=0.5, size=14),
        axis.title.y = element_text(face="bold", size=12),
        axis.text.y = element_text(size=12))

# Plotting yearly Values of GPP_EC
# Figure 7: mean yearly values for the subset
F7 <- ggplot(subset_pheno, aes(factor(years_pheno), mean_GPP_DT_gCm2day)) +
  geom_boxplot(fill="#2D708EFF") +
  geom_jitter(width = 0) +
  ggtitle("GPP_EC Phenophase Subset Values by Year") +
  scale_x_discrete(breaks= years_pheno, labels=years_pheno) +
  xlab("Year") +
  ylab(expression(paste("GPP EC - gC/m"^{2} * "/day"))) +
  theme_bw() +
  theme(plot.title = element_text(lineheight=.8, face="bold", size=14),
        legend.position="none",
        axis.title.x = element_text(face="bold", size=12),
```

```
        axis.text.x = element_text(angle=90, vjust=0.5, size=14),
        axis.title.y = element_text(face="bold", size=12),
        axis.text.y = element_text(size=12))

#### Plotting yearly Values of MODIS GPP
#### Figure 8: mean yearly values for the subset
F8 <- ggplot(subset_pheno, aes(factor(years_pheno), MODISGPP_gCm2day)) +
  geom_boxplot(fill="#2D708EFF") +
  geom_jitter(width = 0) +
  ggtitle("MODIS Phenophase Subset Values by Month") +
  scale_x_discrete(breaks= years_pheno, labels=years_pheno) +
  xlab("Year") +
  ylab(expression(paste("MODIS GPP 500m - gC/m"^{2} * "/day"))) +
  theme_bw() +
  theme(plot.title = element_text(lineheight=.8, face="bold", size=14),
        legend.position="none", axis.title.x = element_text(face="bold", size=14),
        axis.text.x = element_text(angle=90, vjust=0.5, size=14),
        axis.title.y = element_text(face="bold", size=12),
        axis.text.y = element_text(size=12))

# Combine all plots into one graphic
combined_plot_2 <- grid.arrange(F5, F6, F7, F8, ncol = 2)

###### Plot Monthly z-scores #####
names(subset_pheno)

### Monthly Plots of fAPAR z-scores#####
# Figure 9: mean monthly z-scores for the subset
F9 <- ggplot(subset_pheno, aes(factor(months_pheno), fAPAR_z)) +
  geom_boxplot(fill="#2D708EFF") +
  geom_jitter(width = 0) +
  ggtitle("fAPAR Phenophase Subset z-scores by Month") +
  scale_x_discrete(breaks= months_pheno, labels=monthAbbrev_pheno) +
  xlab("Month") +
  ylab(expression(paste("fAPAR (z-scores)"))) +
  theme_bw() +
  theme(plot.title = element_text(lineheight=.8, face="bold", size=14),
        legend.position="none",
        axis.title.x = element_text(face="bold", size=12),
        axis.text.x = element_text(angle=90, vjust=0.5, size=14),
        axis.title.y = element_text(face="bold", size=12),
        axis.text.y = element_text(size=12)) +
  guides(fill=FALSE) +
  geom_hline(yintercept = 0, color = "red")+
  annotate("text", x = 8, y = +0.1, label = "Above historical 'normal' growth", color =
"red")+
  annotate("text", x = 8, y = -0.1, label = "Below historical 'normal' growth", color =
"red")

# Plotting Monthly z-scores of GDMP
# Figure 10: mean monthly z-scores for the subset
F10 <- ggplot(subset_pheno, aes(factor(months_pheno), GDMP_z)) +
  geom_boxplot(fill="#2D708EFF") +
  geom_jitter(width = 0) +
  ggtitle("GDMP Phenophase Subset z-scores by Month") +
  scale_x_discrete(breaks= months_pheno, labels=monthAbbrev_pheno) +
  xlab("Month") +
  ylab(expression(paste("GDMP (z-scores)"))) +
  theme_bw() +
```

```
  theme(plot.title = element_text(lineheight=.8, face="bold", size=14),
        legend.position="none",
        axis.title.x = element_text(face="bold", size=12),
        axis.text.x = element_text(angle=90, vjust=0.5, size=14),
        axis.title.y = element_text(face="bold", size=12),
        axis.text.y = element_text(size=12)) +
  guides(fill=FALSE) +
  geom_hline(yintercept = 0, color = "red")+
  annotate("text", x = 7, y = +0.1, label = "Above historical 'normal' growth", color =
"red")+
  annotate("text", x = 7, y = -0.1, label = "Below historical 'normal' growth", color =
"red")

# Plotting Monthly z-scores of GPP_EC z-scores
# Figure 11: mean monthly z-scores for the subset
F11 <- ggplot(subset_pheno, aes(factor(months_pheno), GPP_EC_z)) +
  geom_boxplot(fill="#2D708EFF") +
  geom_jitter(width = 0) +
  ggtitle("GPP_EC Phenophase Subset z-scores by Month") +
  scale_x_discrete(breaks= months_pheno, labels=monthAbbrev_pheno) +
  xlab("Month") +
  ylab(expression(paste("GPP_EC (z-scores)"))) +
  theme_bw() +
  theme(plot.title = element_text(lineheight=.8, face="bold", size=14),
        legend.position="none",
        axis.title.x = element_text(face="bold", size=12),
        axis.text.x = element_text(angle=90, vjust=0.5, size=14),
        axis.title.y = element_text(face="bold", size=12),
        axis.text.y = element_text(size=12)) +
  guides(fill=FALSE)+
  geom_hline(yintercept = 0, color = "red")+
  annotate("text", x = 8, y = +0.1, label = "Above historical 'normal' growth", color =
"red")+
  annotate("text", x = 8, y = -0.1, label = "Below historical 'normal' growth", color =
"red")

#### Plotting Monthly z-scores of MODIS GPP
#### Figure 12: mean monthly z-scores for the subset
F12 <- ggplot(subset_pheno, aes(factor(months_pheno), MODIS_GPP_z)) +
  geom_boxplot(fill="#2D708EFF") +
  geom_jitter(width = 0) +
  ggtitle("MODIS Phenophase Subset z-scores by Month") +
  scale_x_discrete(breaks= months_pheno, labels=monthAbbrev_pheno) +
  xlab("Month") +
  ylab(expression(paste("MODIS GPP 500m (z-scores)"))) +
  theme_bw() +
  theme(plot.title = element_text(lineheight=.8, face="bold", size=14),
        legend.position="none", axis.title.x = element_text(face="bold", size=14),
        axis.text.x = element_text(angle=90, vjust=0.5, size=14),
        axis.title.y = element_text(face="bold", size=12),
        axis.text.y = element_text(size=12)) +
  guides(fill=FALSE)+
  geom_hline(yintercept = 0, color = "red")+
  annotate("text", x = 6, y = +0.1, label = "Above historical 'normal' growth", color =
"red")+
  annotate("text", x = 6, y = -0.1, label = "Below historical 'normal' growth", color =
"red")

# Combine all plots into one graphic
```

```
combined_plot_3 <- grid.arrange(F9, F10, F11, F12, ncol = 2)

### Yearly Plots ########
# Plotting yearly z-scores of fAPAR
# Figure 13: mean yearly z-scores for the subset
F13 <- ggplot(subset_pheno, aes(factor(years_pheno), fAPAR_z)) +
  geom_boxplot(fill="#2D708EFF") +
  geom_jitter(width = 0) +
  ggtitle("fAPAR Phenophase Subset z-scores by Year") +
  scale_x_discrete(breaks= years_pheno, labels=years_pheno) +
  xlab("Year") +
  ylab(expression(paste("fAPAR (z-scores) "))) +
  theme_bw() +
  theme(plot.title = element_text(lineheight=.8, face="bold", size=14),
        legend.position="none",
        axis.title.x = element_text(face="bold", size=12),
        axis.text.x = element_text(angle=90, vjust=0.5, size=14),
        axis.title.y = element_text(face="bold", size=12),
        axis.text.y = element_text(size=12)) +
  guides(fill=FALSE)+
  geom_hline(yintercept = 0, color = "red")+
  annotate("text", x = 2, y = +0.1, label = "Above historical 'normal' growth", color =
"red")+
  annotate("text", x = 2, y = -0.1, label = "Below historical 'normal' growth", color =
"red")

# Plotting yearly z_scores of GDMP
# Figure 14: mean yearly z-scores for the subset
F14 <- ggplot(subset_pheno, aes(factor(years_pheno), GDMP_z)) +
  geom_boxplot(fill="#2D708EFF") +
  geom_jitter(width = 0) +
  ggtitle("GDMP Phenophase Subset z-scores by Year") +
  scale_x_discrete(breaks= years_pheno, labels=years_pheno) +
  xlab("Year") +
  ylab(expression(paste("GDMP (z-scores)"))) +
  theme_bw() +
  theme(plot.title = element_text(lineheight=.8, face="bold", size=14),
        legend.position="none",
        axis.title.x = element_text(face="bold", size=12),
        axis.text.x = element_text(angle=90, vjust=0.5, size=14),
        axis.title.y = element_text(face="bold", size=12),
        axis.text.y = element_text(size=12)) +
  guides(fill=FALSE)+
  geom_hline(yintercept = 0, color = "red")+
  annotate("text", x = 2, y = +1, label = "Above historical 'normal' growth", color = "red")+
  annotate("text", x = 2, y = -1.4, label = "Below historical 'normal' growth", color =
"red")

# Plotting yearly z-scores of GPP_EC
# Figure 15: mean yearly values for the subset
F15 <- ggplot(subset_pheno, aes(factor(years_pheno), GPP_EC_z)) +
  geom_boxplot(fill="#2D708EFF") +
  geom_jitter(width = 0) +
  ggtitle("GPP_EC Phenophase Subset z-scores by Year") +
  scale_x_discrete(breaks= years_pheno, labels=years_pheno) +
  xlab("Year") +
  ylab(expression(paste("GPP_EC (z-scores)"))) +
  theme_bw() +
  theme(plot.title = element_text(lineheight=.8, face="bold", size=14),
```

```
        legend.position="none",
        axis.title.x = element_text(face="bold", size=12),
        axis.text.x = element_text(angle=90, vjust=0.5, size=14),
        axis.title.y = element_text(face="bold", size=12),
        axis.text.y = element_text(size=12)) +
  guides(fill=FALSE)+
  geom_hline(yintercept = 0, color = "red")+
  annotate("text", x = 2, y = +0.1, label = "Above historical 'normal' growth", color =
"red")+
  annotate("text", x = 2, y = -0.1, label = "Below historical 'normal' growth", color =
"red")

#### Plotting yearly z-scores of MODIS GPP
#### Figure 16: mean yearly z-scores for the subset
F16 <- ggplot(subset_pheno, aes(factor(years_pheno), MODIS_GPP_z)) +
  geom_boxplot(fill="#2D708EFF") +
  geom_jitter(width = 0) +
  ggtitle("MODIS Phenophase Subset z-scores by Month") +
  scale_x_discrete(breaks= years_pheno, labels=years_pheno) +
  xlab("Year") +
  ylab(expression(paste("MODIS GPP (z-scores)"))) +
  theme_bw() +
  theme(plot.title = element_text(lineheight=.8, face="bold", size=14),
        legend.position="none", axis.title.x = element_text(face="bold", size=14),
        axis.text.x = element_text(angle=90, vjust=0.5, size=14),
        axis.title.y = element_text(face="bold", size=12),
        axis.text.y = element_text(size=12)) +
  guides(fill=FALSE)+
  geom_hline(yintercept = 0, color = "red")+
  annotate("text", x = 2, y = +1.5, label = "Above historical 'normal' growth", color =
"red")+
  annotate("text", x = 2, y = -1.6, label = "Below historical 'normal' growth", color =
"red")

# Combine all plots into one graphic
combined_plot_4 <- grid.arrange(F13, F14, F15, F16, ncol = 2)

save.image("G:/AAA.Google_Drive/THESIS_BIOMASS_ANOMALIES/R_Processing/Pheno_sub_data_sub_NOVE
MBER.ANALYSIS.RData")

##############################

### Linearity & Correlation Plots ###
library(ggplot2)
library(ggpubr)

## fAPAR(z-scores) : GDMP (z-scores) ##
x1 = subset_pheno$fAPAR_z
y1 = subset_pheno$GDMP_z
pair1 <- data.frame(x1,y1)
colnames(pair1)
colnames(pair1) <- c("fAPAR", "GDMP")

ggscatter(pair1, x = "fAPAR", y = "GDMP",
          add = "reg.line",                                 # Add regression line
          conf.int = TRUE,                                  # Add confidence interval
          cor.coef.coord = c(1, 1),
          add.params = list(color = "blue",
                            fill = "lightgray")
```

```
)+
  stat_cor(method = "pearson", label.x = 1, label.y = -0.5) +  # Add correlation coefficient
  xlab("fAPAR (z-scores)") +
  ylab("GDMP (z-scores)")

## GPPEC (z-scores) : fAPAR (z-scores) ##
x2 = subset_pheno$fAPAR_z
y2 = subset_pheno$GPP_EC_z

pair2 <- data.frame(x2,y2)
colnames(pair2)
colnames(pair2) <- c("fAPAR", "GPPEC")

ggscatter(pair2, x = "fAPAR", y = "GPPEC",
          add = "reg.line",                                 # Add regression line
          conf.int = TRUE,                                  # Add confidence interval
          cor.coef.coord = c(1, 1),
          add.params = list(color = "blue",
                            fill = "lightgray")
)+
  stat_cor(method = "pearson", label.x = 1, label.y = 1) +  # Add correlation coefficient
  xlab("fAPAR (z-scores)") +
  ylab("GPP_EC (z-scores)")

## GPPEC (z-scores) : GDMP (z-scores) ##
x3 = subset_pheno$GDMP_z
y3 = subset_pheno$GPP_EC_z

pair3 <- data.frame(x3,y3)
colnames(pair3)
colnames(pair3) <- c("GDMP", "GPP_EC")

ggscatter(pair3, x = "GDMP", y = "GPP_EC",
          add = "reg.line",                                 # Add regression line
          conf.int = TRUE,                                  # Add confidence interval
          cor.coef.coord = c(1, 1),
          add.params = list(color = "blue",
                            fill = "lightgray")
)+
  stat_cor(method = "pearson", label.x = 1.5, label.y = 0) + # Add correlation coefficient
  xlab("GDMP (z-scores)") +
  ylab("GPP_EC (z-scores)")

## MODIS GPP (z-scores) : fAPAR (z-scores) ##
x4 = subset_pheno$fAPAR_z
y4 = subset_pheno$MODIS_GPP_z

pair4 <- data.frame(x4,y4)
colnames(pair4)
colnames(pair4) <- c("fAPAR", "MODIS_GPP")

ggscatter(pair4, x = "fAPAR", y = "MODIS_GPP",
          add = "reg.line",                                 # Add regression line
          conf.int = TRUE,                                  # Add confidence interval
          cor.coef.coord = c(1, 1),
          add.params = list(color = "blue",
                            fill = "lightgray")
)+
  stat_cor(method = "pearson", label.x = -0.5, label.y = 2) +  # Add correlation coefficient
```

```
  xlab("fAPAR (z-scores)") +
  ylab("MODIS GPP (z-scores)")


## MODIS GPP (z-scores) : GDMP (z-scores) ##
x5 = subset_pheno$GDMP_z
y5 = subset_pheno$MODIS_GPP_z

pair5 <- data.frame(x5,y5)
colnames(pair5)
colnames(pair5) <- c("GDMP", "MODIS_GPP")

ggscatter(pair5, x = "GDMP", y = "MODIS_GPP",
          add = "reg.line",                                 # Add regression line
          conf.int = TRUE,                                  # Add confidence interval
          cor.coef.coord = c(1, 1),
          add.params = list(color = "blue",
                            fill = "lightgray")
)+
  stat_cor(method = "pearson", label.x = -0.5, label.y = 2) +  # Add correlation coefficient
  xlab("GDMP (z-scores)") +
  ylab("MODIS GPP (z-scores)")

## MODIS GPP (z-scores) : GPP_EC (z-scores) ##
x6 = subset_pheno$GPP_EC_z
y6 = subset_pheno$MODIS_GPP_z

pair6 <- data.frame(x6,y6)
colnames(pair6)
colnames(pair6) <- c("GPP_EC", "MODIS_GPP")

ggscatter(pair6, x = "GPP_EC", y = "MODIS_GPP",
          add = "reg.line",                                 # Add regression line
          conf.int = TRUE,                                  # Add confidence interval
          cor.coef.coord = c(1, 1),
          add.params = list(color = "blue",
                            fill = "lightgray")
)+
  stat_cor(method = "pearson", label.x = -0.5, label.y = 2) +  # Add correlation coefficient
  xlab("GPP_EC (z-scores)") +
  ylab("MODIS GPP (z-scores)")

##### Seasonal Analysis ######
# Create subsets for each year
subset_2014_season <- subset(subset_pheno, Date >= as.Date("2013-10-26") & Date <=
as.Date("2014-05-29"))
subset_2015_season <- subset(subset_pheno, Date >= as.Date("2014-11-06") & Date <=
as.Date("2015-06-18"))
subset_2016_season <- subset(subset_pheno, Date >= as.Date("2015-01-08") & Date <=
as.Date("2016-05-15"))
subset_2017_season <- subset(subset_pheno, Date >= as.Date("2016-12-24") & Date <=
as.Date("2017-05-16"))
subset_2018_season <- subset(subset_pheno, Date >= as.Date("2018-01-13") & Date <=
as.Date("2018-06-16"))
subset_2019_season <- subset(subset_pheno, Date >= as.Date("2019-01-15") & Date <=
as.Date("2019-05-31"))
subset_2020_season <- subset(subset_pheno, Date >= as.Date("2020-01-01") & Date <=
as.Date("2020-05-18"))
```

```
subset_2021_season <- subset(subset_pheno, Date >= as.Date("2020-12-18") & Date <=
as.Date("2021-04-22"))

names(subset_2014_season)

# Descriptive Statistics of z-scores ### AI Assisted loop using ChatGPT
seasons <- c(2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021)

### fAPAR ###
for (season in seasons) {
  subset <- get(paste0("subset_", season, "_season"))
  fAPAR_z <- subset$fAPAR_z

  cat("Season:", season, "\n")
  cat("Summary:", round(summary(fAPAR_z, na.rm = TRUE),3), "\n")
  cat("Standard Deviation:", round(sd(fAPAR_z, na.rm = TRUE),3), "\n")
  cat("n=", sum(!is.na(subset$fAPAR_z)), "\n")
  cat("Sum:" , sum(fAPAR_z, na.rm = TRUE), "\n")
  cat("Mean:" , mean(fAPAR_z, na.rm = TRUE), "\n")
}

### GDMP ###
for (season in seasons) {
  subset <- get(paste0("subset_", season, "_season"))
  GDMP_z <- subset$GDMP_z

  cat("Season:", season, "\n")
  cat("Summary:", round(summary(GDMP_z, na.rm = TRUE),3), "\n")
  cat("Standard Deviation:", round(sd(GDMP_z, na.rm = TRUE),3), "\n")
  cat("n=", sum(!is.na(subset$GDMP_z)), "\n")
  cat("Sum:" , sum(GDMP_z, na.rm = TRUE), "\n")
  cat("Mean:" , mean(GDMP_z, na.rm = TRUE), "\n")
}

### GPP EC ###
for (season in seasons) {
  subset <- get(paste0("subset_", season, "_season"))
  GPP_EC_z <- subset$GPP_EC_z

  cat("Season:", season, "\n")
  cat("Summary:", round(summary(GPP_EC_z, na.rm = TRUE),3), "\n")
  cat("Standard Deviation:", round(sd(GPP_EC_z, na.rm = TRUE),3), "\n")
  cat("n=", sum(!is.na(subset$GPP_EC_z)), "\n")
  cat("Sum:" , sum(GPP_EC_z, na.rm = TRUE), "\n")
  cat("Mean:" , mean(GPP_EC_z, na.rm = TRUE), "\n")
}

### MODIS_GPP_z ###
for (season in seasons) {
  subset <- get(paste0("subset_", season, "_season"))
  MODIS_GPP_z <- subset$MODIS_GPP_z

  cat("Season:", season, "\n")
  cat("Summary:", round(summary(MODIS_GPP_z, na.rm = TRUE),3), "\n")
  cat("Standard Deviation:", round(sd(MODIS_GPP_z, na.rm = TRUE),3), "\n")
  cat("n=", sum(!is.na(subset$MODIS_GPP_z)), "\n")
  cat("Sum:" , sum(MODIS_GPP_z, na.rm = TRUE), "\n")
  cat("Mean:" , mean(MODIS_GPP_z, na.rm = TRUE), "\n")
}
```

```
#### Test for Normality
variables <- data.frame(
  fAPAR_unitless = ifelse(is.na(subset_pheno$fAPAR_unitless), NA,
as.numeric(subset_pheno$fAPAR_unitless)),
  GDMP = ifelse(is.na(subset_pheno$GDMP), NA, as.numeric(subset_pheno$GDMP)),
  GPP_EC = ifelse(is.na(subset_pheno$mean_GPP_DT_gCm2day), NA,
as.numeric(subset_pheno$mean_GPP_DT_gCm2day)),
  MODIS_GPP = ifelse(is.na(subset_pheno$MODISGPP_gCm2day), NA,
as.numeric(subset_pheno$MODISGPP_gCm2day))
)

# Iterate over column names of variables dataframe
for (variable in colnames(variables)) {
  data <- na.omit(variables[[variable]]) # Remove NAs from the column
  numeric_data <- as.numeric(data, na.rm = TRUE) # Convert column to numeric type

  result <- shapiro.test(numeric_data) # Perform Shapiro-Wilk test on the column
  print(paste("Variable:", variable)) # Print variable name
  print(result) # Print test result
}

########################
### Test for Anomalous Artefacts / Outliers ###
########################

### fAPAR ###
# Calculate mean and standard deviation
mean_fAPAR <- mean(subset_pheno$fAPAR_unitless, na.rm = TRUE)
sd_fAPAR <- sd(subset_pheno$fAPAR_unitless, na.rm = TRUE)
print(sum(!is.na(subset_pheno$fAPAR_unitless)))

# Create the box plot
boxplot(subset_pheno$fAPAR_unitless, ylim = range(0.6, -0.6), main = "Boxplot of fAPAR
Dataset", ylab = "fAPAR (Unitless)")
text(1, mean_fAPAR, paste("Mean:", round(mean_fAPAR, 2)), pos = 3)
text(1, mean_fAPAR - 1.2 * sd_fAPAR, paste("SD:", round(sd_fAPAR, 2)), pos = 3)

### GDMP ###
# Calculate mean and standard deviation
mean_GDMP <- mean(subset_pheno$GDMP, na.rm = TRUE)
sd_GDMP <- sd(subset_pheno$GDMP, na.rm = TRUE)
print(sum(!is.na(subset_pheno$GDMP)))

# Create the box plot
boxplot(subset_pheno$GDMP, ylim = range(8000, -5500), main = "Boxplot of GDMP Dataset", ylab
= "GDMP (kgDM/ha/day)")
text(1, mean_GDMP + 100.0, paste("Mean:", round(mean_GDMP, 2)), pos = 2)
text(1, mean_GDMP - 1.2 * sd_GDMP, paste("SD:", round(sd_GDMP, 2)), pos = 2)

### GPP_EC ###
# Calculate mean and standard deviation
mean_GPP_EC <- mean(subset_pheno$mean_GPP_DT_gCm2day, na.rm = TRUE)
sd_GPP_EC <- sd(subset_pheno$mean_GPP_DT_gCm2day, na.rm = TRUE)
print(sum(!is.na(subset_pheno$mean_GPP_DT_gCm2day)))

# Create the boxplot
boxplot(subset_pheno$mean_GPP_DT_gCm2day, ylim = range(7, -5), main = "Boxplot of GPP_EC
Dataset", ylab = "GPP_EC (gC/m2/day)")
```

```
text(1, mean_GPP_EC, paste("Mean:", round(mean_GPP_EC, 2)), pos = 2)
text(1, mean_GPP_EC - 1.5 * sd_GPP_EC, paste("SD:", round(sd_GPP_EC, 2)), pos = 2)

#### MODIS GPP ####
# Calculate mean and standard deviation
mean_MODIS_GPP <- mean(subset_pheno$MODISGPP_gCm2day, na.rm = TRUE)
sd_MODIS_GPP <- sd(subset_pheno$MODISGPP_gCm2day, na.rm = TRUE)
print(sum(!is.na(subset_pheno$MODISGPP_gCm2day)))

# Create the boxplot
boxplot(subset_pheno$MODISGPP_gCm2day, ylim = range(4.0, -4.0), main = "Boxplot of MODIS GPP
Dataset", ylab = "MODIS GPP (gC/m2/day)")
text(1, mean_MODIS_GPP + 0.2, paste("Mean:", round(mean_MODIS_GPP, 2)), pos = 2)
text(1, mean_MODIS_GPP - 1.3 * sd_MODIS_GPP, paste("SD:", round(sd_MODIS_GPP, 2)), pos = 2)

### Test for Skewness - The larger the value of skewness, the larger the distribution differs
from a normal distribution.###
library(moments)
## fAPAR
fAPAR_skew <- skewness(subset_pheno$fAPAR_unitless, na.rm = TRUE)
par(mfrow=c(1,1))
hist(subset_pheno$fAPAR_unitless, freq = FALSE, main = "Histogram of fAPAR with Skewness and
Normal Curve", xlab = "fAPAR (Unitless)", ylim = c(0,6))
curve(dnorm(x, mean = mean(subset_pheno$fAPAR_unitless, na.rm = TRUE), sd =
sd(subset_pheno$fAPAR_unitless, na.rm = TRUE)), col = "blue", lwd = 2, add = TRUE)
text(0.2, 5, paste0("Skewness = ", round(fAPAR_skew, 2)))

## GDMP
GDMP_skew <- skewness(subset_pheno$GDMP, na.rm = TRUE)
par(mfrow=c(1,1))
hist(subset_pheno$GDMP, freq = FALSE, main = "Histogram of GDMP with Skewness and Normal
Curve", xlab = "GDMP (kgDM/ha/day)", ylim = c(0,0.0006))
curve(dnorm(x, mean = mean(subset_pheno$GDMP, na.rm = TRUE), sd = sd(subset_pheno$GDMP, na.rm
= TRUE)), col = "blue", lwd = 2, add = TRUE)
text(2500, 0.0005, paste0("Skewness = ", round(GDMP_skew, 2)))

## GPP_EC
GPP_EC_skew <- skewness(subset_pheno$mean_GPP_DT_gCm2day, na.rm = TRUE)
par(mfrow=c(1,1))
hist(subset_pheno$mean_GPP_DT_gCm2day, freq = FALSE, main = "Histogram of Eddy Covariance GPP
with Skewness and Normal Curve", xlab = "GPP EC (gC/m2/day)", ylim = c(0,0.6))
curve(dnorm(x, mean = mean(subset_pheno$mean_GPP_DT_gCm2day, na.rm = TRUE), sd =
sd(subset_pheno$mean_GPP_DT_gCm2day, na.rm = TRUE)), col = "blue", lwd = 2, add = TRUE)
text(1.5, 0.5, paste0("Skewness = ", round(GPP_EC_skew, 2)))

## MODIS GPP
MODIS_GPP_skew <- skewness(subset_pheno$MODIS_mean_kgCm2_8d, na.rm = TRUE)
par(mfrow=c(1,1))
hist(subset_pheno$MODIS_mean_kgCm2_8d, freq = FALSE, main = "Histogram of MODIS GPP with
Skewness and Normal Curve", xlab = "MODIS GPP (kgC/m2/day)", ylim = c(0,70.0))
curve(dnorm(x, mean = mean(subset_pheno$MODIS_mean_kgCm2_8d, na.rm = TRUE), sd =
sd(subset_pheno$MODIS_mean_kgCm2_8d, na.rm = TRUE)), col = "blue", lwd = 2, add = TRUE)
text(0.023, 60, paste0("Skewness = ", round(MODIS_GPP_skew, 2)))

# Shapiro-Wilk test for normality - If the value of the Shapiro-Wilk Test is greater than
0.05, the data is normal
result_fAPAR <- shapiro.test(subset_pheno$fAPAR_unitless)
print(result_fAPAR)
result_GDMP <- shapiro.test(subset_pheno$GDMP)
```

```
print(result_GDMP)
result_GPP_EC <- shapiro.test(subset_pheno$GPP_EC)
print(result_GPP_EC)
result_MODIS_GPP <- shapiro.test(subset_pheno$MODIS_GPP)
print(result_MODIS_GPP)

### LINEAR REGRESSION ###
#### https://towardsdatascience.com/understanding-linear-regression-output-in-r-7a9cbda948b3
###
library(dplyr)
library(caret)

### Linear Model - GPP_EC  ~ fAPAR  ###
#### Split the data into Model Training and Model Testing Data
set.seed(123)
fAPAR_training_index <- subset_pheno$fAPAR_unitless
fAPAR_training_index <- na.omit(fAPAR_training_index)
complete_cases_fAPAR <- complete.cases(subset_pheno$fAPAR)
subset_pheno_fAPAR <- subset_pheno[complete_cases_fAPAR, ]
fAPAR_training_index <- createDataPartition(fAPAR_training_index, p = 0.6, list = FALSE)
fAPAR_training_data <- subset_pheno_fAPAR[fAPAR_training_index, ]
fAPAR_test_data <- subset_pheno_fAPAR[-fAPAR_training_index, ]

### Train fAPAR Model ###
model_fAPAR <- lm(mean_GPP_DT_gCm2day  ~ fAPAR_unitless, data = fAPAR_training_data)
summary(model_fAPAR)

#### Predict GPP Values using the fAPAR Model
fAPAR_model_predictions <- as.data.frame(predict(model_fAPAR, newdata = fAPAR_test_data,
interval = 'confidence', level = 0.95))

# Calculate the fAPAR model accuracy RMSE
fAPAR_accuracy <- sqrt(mean((fAPAR_test_data$mean_GPP_DT_gCm2day -
fAPAR_model_predictions$fit)^2))
#print(paste("Accuracy:", round(fAPAR_accuracy, 2)))
print(paste("Root Mean Square Error (gC/m2/day):", round(RMSE(fAPAR_model_predictions$fit,
fAPAR_test_data$mean_GPP_DT_gCm2day), 2)))

# Load the ggplot2 library
library(ggplot2)

# Create the plot
ggplot(data = fAPAR_model_predictions) +
  geom_line(aes(x = seq_along(fit), y = fit), color = "blue") +
  geom_ribbon(aes(x = seq_along(fit), ymin = lwr, ymax = upr), fill = "gray", alpha = 0.4) +
  geom_text(aes(label = "CI = 95%", x = 12, y = 3), color = "black") +
  labs(x = "Observation No.", y = "Predicted GPP (gC/m2/day)", title = "GPP Predicted from
fAPAR with 95% Confidence Intervals (Greyed Area)") +
  theme_minimal() +
  theme(axis.text = element_text(size = 16),
        axis.title = element_text(size = 18),
        plot.title = element_text(size = 18))

length(fAPAR_model_predictions$fit)
length(model_fAPAR$resid)
length(fAPAR_training_data$mean_GPP_DT_gCm2day)

### Plot Residuals - fAPAR #
# fAPAR x Residuals Plot
```

```
plot(model_fAPAR$resid ~
fAPAR_training_data$mean_GPP_DT_gCm2day[order(fAPAR_training_data$mean_GPP_DT_gCm2day)],
     main = expression("Panel A. GPP ("~gC/m^2/day*") - Residuals for Linear Regression"),
     xlab = expression("Training GPP ("~gC/m^2/day*")"),
     ylab = expression("Residuals ("~gC/m^2/day*")"),
     cex.lab = 1.0,
     xlim = range(model_fAPAR$resid),
     xaxt = "n" # Remove x-axis labels)
)
abline(h = 0, lty = 2, col = "red")
axis(1, at = pretty(fAPAR_training_data$mean_GPP_DT_gCm2day), labels =
pretty(fAPAR_training_data$mean_GPP_DT_gCm2day), cex.axis = 0.7)

# Histogram of Residuals
hist(model_fAPAR$resid, main = "Panel B. Histogram of Residuals",
     xlab = expression("Training GPP ("~gC/m^2/day*")"),
     ylab = expression("Residuals ("~gC/m^2/day*")"), cex.lab = 1.0,
     breaks = 30)

# Q-Q Plot with adjusted axis limits
qqnorm(model_fAPAR$resid, main = "Panel C. Normal Q-Q Plot", cex.lab = 1.0,
       xlim = c(-4, 4), ylim = c(-4, 4))
abline(a = 0, b = 1, col = "red")  # 45-degree reference line
text(1.5, 0.8, "1:1 Reference Line", pos = 3, col = "red")

### Linear Model - GDMP ###
#### Split the data into Model Training and Model Testing Data
library(dplyr)
library(caret)
set.seed(123)
GDMP_training_index <- na.omit(subset_pheno$mean_GPP_DT_gCm2day)
GDMP_training_index <- na.omit(GDMP_training_index)
complete_cases_GDMP <- complete.cases(subset_pheno$mean_GPP_DT_gCm2day)
subset_pheno_GDMP <- subset_pheno[complete_cases_GDMP, ]
GDMP_training_index <- createDataPartition(GDMP_training_index, p = 0.6, list = FALSE)
GDMP_training_data <- subset_pheno_GDMP[GDMP_training_index, ]
GDMP_test_data <- subset_pheno_GDMP[-GDMP_training_index, ]

### Train GDMP Model ###
model_GDMP <- lm(mean_GPP_DT_gCm2day ~ GDMP, data = GDMP_training_data)
summary(model_GDMP)

#### Predict GPP Values using the GDMP Model
GDMP_model_predictions <- as.data.frame(predict(model_GDMP, newdata = GDMP_test_data,
interval = 'confidence', level = 0.95))

length(GDMP_model_predictions$fit)
length(model_GDMP$resid)
length(GDMP_training_data$mean_GPP_DT_gCm2day)

# Create the plot
ggplot(data = GDMP_model_predictions) +
  geom_line(aes(x = seq_along(fit), y = fit), color = "blue") +
  geom_ribbon(aes(x = seq_along(fit), ymin = lwr, ymax = upr), fill = "gray", alpha = 0.4) +
  geom_text(aes(label = "CI = 95%", x = 12.5, y = 2.8), color = "black") +
  labs(x = "Observation No.", y = "Predicted GPP Value (gC/m2/day)", title = "GPP Predicted
from GDMP with 95% Confidence Intervals (Greyed Area)") +
  theme_minimal() +
  theme(axis.text = element_text(size = 16),
```

```
        axis.title = element_text(size = 18),
        plot.title = element_text(size = 18))

# Calculate the GDMP model accuracy RMSE
GDMP_accuracy <- sqrt(mean((GDMP_test_data$mean_GPP_DT_gCm2day -
GDMP_model_predictions$fit)^2, na.rm = TRUE))
print(paste("Accuracy:", round(GDMP_accuracy, 3)))
print(paste("Root Mean Square Error (gC/m2/day):", round(RMSE(GDMP_model_predictions$fit,
GDMP_test_data$mean_GPP_DT_gCm2day), 3)))

### Plot Residuals - GDMP #
# GDMP x Residuals Plot
plot(model_GDMP$resid ~
GDMP_training_data$mean_GPP_DT_gCm2day[order(GDMP_training_data$mean_GPP_DT_gCm2day)],
     main=expression("Panel A. GPP ("~gC/m^2/day*") - Residuals for Linear Regression"),
     xlab = expression("Training GPP ("~gC/m^2/day*")"),
     ylab = expression("Residuals ("~gC/m^2/day*")"),
     cex.lab = 1.0,
     xlim = range(model_GDMP$resid),
     xaxt = "n" # Remove x-axis labels)
)
abline(h=0,lty=2, col = "red")
axis(1, at = pretty(GDMP_training_data$mean_GPP_DT_gCm2day), labels =
pretty(GDMP_training_data$mean_GPP_DT_gCm2day), cex.axis = 1.0)

#Histogram of Residuals
hist(model_GDMP$resid, main="Panle B. Histogram of Residuals",
     xlab = expression("Training GPP ("~gC/m^2/day*")"),
     ylab = expression("Residuals ("~gC/m^2/day*")"), cex.lab = 1.0,
     breaks = 30)

#Q-Q Plot
qqnorm(model_GDMP$resid, main = "Panel C. Normal Q-Q Plot", cex.lab = 1.0,
       xlim = c(-4, 4), ylim = c(-4, 4))
abline(a = 0, b = 1, col = "red")  # 45-degree reference line
text(1.5, 2.0, "1:1 Reference Line", pos = 3, col = "red")

### Linear Model - MODIS_GPP ###
#### Split the data into Model Training and Model Testing Data
set.seed(123)
MODISGPP_training_index <- subset_pheno$mean_GPP_DT_gCm2day
MODISGPP_training_index <- na.omit(MODISGPP_training_index)
complete_cases_MODISGPP <- complete.cases(subset_pheno$mean_GPP_DT_gCm2day)
subset_pheno_MODISGPP <- subset_pheno[complete_cases_MODISGPP, ]
MODISGPP_training_index <- createDataPartition(MODISGPP_training_index, p = 0.6, list =
FALSE)
MODISGPP_training_data <- subset_pheno_MODISGPP[MODISGPP_training_index, ]
MODISGPP_test_data <- subset_pheno_MODISGPP[-MODISGPP_training_index, ]

### Train MODIS_GPP Model ###
model_MODIS_GPP <- lm(mean_GPP_DT_gCm2day ~ MODISGPP_gCm2day, data = MODISGPP_training_data)
summary(model_MODIS_GPP)

# Assuming 'model_MODIS_GPP' holds your linear regression model
summary_info <- summary(model_MODIS_GPP)
num_observations <- summary_info$df[1] + summary_info$df[2]  # Total degrees of freedom
(numerator + denominator)
cat("Number of observation pairs:", num_observations, "\n")
```

```
#### Predict GPP Values using the MODIS_GPP Model
MODIS_GPP_model_predictions <- as.data.frame(predict(model_MODIS_GPP, newdata =
MODISGPP_test_data, interval = 'confidence', level = 0.95))
length(MODIS_GPP_model_predictions$fit)

# Calculate the MODIS model accuracy RMSE
MODIS_accuracy <- sqrt(mean((MODISGPP_test_data$MODISGPP_gCm2day -
MODIS_GPP_model_predictions$fit)^2, na.rm = TRUE))
print(paste("Accuracy:", round(MODIS_accuracy, 3)))
print(paste("Root Mean Square Error (gC/m2/day):",
round(RMSE(MODIS_GPP_model_predictions$fit, MODISGPP_test_data$MODISGPP_gCm2day), 3)))

# Create the plot
ggplot(data = MODIS_GPP_model_predictions) +
  geom_line(aes(x = seq_along(fit), y = fit), color = "blue") +
  geom_ribbon(aes(x = seq_along(fit), ymin = lwr, ymax = upr), fill = "gray", alpha = 0.4) +
  geom_text(aes(label = "CI = 95%", x = 16.0, y = 2.7), color = "black") +
  labs(x = "Observation No.", y = "Predicted GPP Value (gC/m2/day)", title = "GPP Predicted
from MODIS GPP with 95% Confidence Intervals (Greyed Area)") +
  theme_minimal() +
  theme(axis.text = element_text(size = 16),
        axis.title = element_text(size = 18),
        plot.title = element_text(size = 18))

### Plot Residuals - MODIS GPPP #
# MODIS_GPP x Residuals Plot
plot(model_MODIS_GPP$resid ~
na.omit(MODISGPP_training_data$mean_GPP_DT_gCm2day[order(MODISGPP_training_data$mean_GPP_DT_g
Cm2day)]),
     main = expression("Panel A. GPP ("~gC/m^2/day*") - Residuals for Linear Regression"),
     xlab = expression("Training GPP ("~gC/m^2/day*")"),
     ylab = expression("Residuals ("~gC/m^2/day*")"),
     cex.lab = 1.0,
     xlim = range(model_MODIS_GPP$resid),
     xaxt = "n" # Remove x-axis labels)
)
abline(h=0,lty=2, col = "red")
axis(1, at = pretty(MODIS_GPP_training_data$mean_GPP_DT_gCm2day), labels =
pretty(MODIS_GPP_training_data$mean_GPP_DT_gCm2day), cex.axis = 1.0)

#Histogram of Residuals
hist(model_MODIS_GPP$resid, main="Panel B. Histogram of Residuals",
     xlab = expression("Training GPP ("~gC/m^2/day*")"),
     ylab = expression("Residuals ("~gC/m^2/day*")"), cex.lab = 1.0,
     breaks = 30)

#Q-Q Plot
qqnorm(model_MODIS_GPP$resid, main = "Panel C. Normal Q-Q Plot", cex.lab = 1.0,
       xlim = c(-4, 4), ylim = c(-4, 4))
abline(a = 0, b = 1, col = "red")  # 45-degree reference line
text(2.2, 2.2, "1:1 Reference Line", pos = 3, col = "red")

### Bar plots ###
par(mfrow = c(4, 1))
par(mar = c(4, 4, 2, 2))

fAPAR_dates <- as.Date(c(subset_pheno$Date))
fAPAR_values <- c(subset_pheno$fAPAR_z)
fAPAR_years <- format(fAPAR_dates, "%Y")
```

```
GDMP_dates <- as.Date(c(subset_pheno$Date))
GDMP_values <- c(subset_pheno$GDMP_z)
GDMP_years <- format(GDMP_dates, "%Y")

GPP_EC_dates <- as.Date(c(subset_pheno$Date))
GPP_EC_values <- c(subset_pheno$GPP_EC_z)
GPP_EC_years <- format(GPP_EC_dates, "%Y")

MODIS_GPP_dates <- as.Date(c(subset_pheno$Date))
MODIS_GPP_values <- c(subset_pheno$MODIS_GPP_z)
MODIS_GPP_years <- format(MODIS_GPP_dates, "%Y")

# Get unique years from fAPAR_years
unique_years <- unique(MODIS_GPP_years)

# Function to generate pastel colors
generate_contrasting_pastel_color <- function(n) {
  hue_values <- seq(15, 375, length.out = n + 1)[-1]  # Varying hue values
  pastel_colors <- matrix(hcl(hue_values, 30, seq(70, 90, length = n)), ncol = 3)
  return(pastel_colors)
}

# Create a custom color palette with a color for each unique year
color_palette <- generate_contrasting_pastel_color(length(unique_years))

# Map colors to years using a named vector
year_color_mapping <- setNames(color_palette, unique_years)

# Use the mapped colors for plotting fAPAR
barplot(fAPAR_values,
        names.arg = fAPAR_dates,
        xlab = "Date",
        ylab = "z-score",
        main = "fAPAR (z-scores) - Phenophase Subset (Greenup-to-Dormancy Only)",
        col = year_color_mapping[fAPAR_years],
        ylim = c(-2.5,2.5))
abline(h = c(1.645, 1.960, -1.645, -1.960), col = c("red", "blue", "red", "blue"))
text(x = 1, y = 1.645 - 0.3, labels = "z = 1.645, 90% Probability Upper Limit", pos = 4, col
= "red")
text(x = 1, y = 1.960 + 0.3, labels = "z = 1.960, 95% Probability Upper Limit", pos = 4, col
= "blue")
text(x = 1, y = -1.645 + 0.3, labels = "z = -1.645, 90% Probability Lower Limit", pos = 4,
col = "red")
text(x = 1, y = -1.960 - 0.3, labels = "z = -1.960, 95% Probability Lower Limit", pos = 4,
col = "blue")

# Use the mapped colors for plotting GDMP
barplot(GDMP_values,
        names.arg = GDMP_dates,
        xlab = "Date",
        ylab = "z-score",
        main = "GDMP (z-scores) - Phenophase Subset (Greenup-to-Dormancy Only)",
        col = year_color_mapping[GDMP_years],
        ylim = c(-2.5,2.5))
abline(h = c(1.645, 1.960, -1.645, -1.960), col = c("red", "blue", "red", "blue"))
text(x = 1, y = 1.645 - 0.3, labels = "z = 1.645, 90% Probability Upper Limit", pos = 4, col
= "red")
```

```
text(x = 1, y = 1.960 + 0.3, labels = "z = 1.960, 95% Probability Upper Limit", pos = 4, col
= "blue")
text(x = 1, y = -1.645 + 0.3, labels = "z = -1.645, 90% Probability Lower Limit", pos = 4,
col = "red")
text(x = 1, y = -1.960 - 0.3, labels = "z = -1.960, 95% Probability Lower Limit", pos = 4,
col = "blue")

# Use the mapped colors for plotting GPP_EC
barplot(GPP_EC_values,
        names.arg = GPP_EC_dates,
        xlab = "Date",
        ylab = "z-score",
        main = "GPP_EC (z-scores) - Phenophase Subset (Greenup-to-Dormancy Only)",
        col = year_color_mapping[GPP_EC_years],
        ylim = c(-2.5,2.5))
abline(h = c(1.645, 1.960, -1.645, -1.960), col = c("red", "blue", "red", "blue"))
text(x = 1, y = 1.645 - 0.3, labels = "z = 1.645, 90% Probability Upper Limit", pos = 4, col
= "red")
text(x = 1, y = 1.960 + 0.3, labels = "z = 1.960, 95% Probability Upper Limit", pos = 4, col
= "blue")
text(x = 1, y = -1.645 + 0.3, labels = "z = -1.645, 90% Probability Lower Limit", pos = 4,
col = "red")
text(x = 1, y = -1.960 - 0.3, labels = "z = -1.960, 95% Probability Lower Limit", pos = 4,
col = "blue")

# Use the mapped colors for plotting MODIS_GPP
barplot(MODIS_GPP_values,
        names.arg = MODIS_GPP_dates,
        xlab = "Date",
        ylab = "z-score",
        main = "MODIS_GPP (z-scores) - Phenophase Subset (Greenup-to-Dormancy Only)",
        col = year_color_mapping[MODIS_GPP_years],
        ylim = c(-2.5,2.5))
abline(h = c(1.645, 1.960, -1.645, -1.960), col = c("red", "blue", "red", "blue"))
text(x = 1, y = 1.645 - 0.3, labels = "z = 1.645, 90% Probability Upper Limit", pos = 4, col
= "red")
text(x = 1, y = 1.960 + 0.3, labels = "z = 1.960, 95% Probability Upper Limit", pos = 4, col
= "blue")
text(x = 1, y = -1.645 + 0.3, labels = "z = -1.645, 90% Probability Lower Limit", pos = 4,
col = "red")
text(x = 1, y = -1.960 - 0.3, labels = "z = -1.960, 95% Probability Lower Limit", pos = 4,
col = "blue")

# Create a legend plot for the color palette
legend_plot <- function(year_color_mapping) {
  n_colors <- length(year_color_mapping)
  legend_width <- 10
  legend_height <- n_colors

  plot(0, 0, type = "n", xlim = c(0, legend_width), ylim = c(0, legend_height),
       xlab = "", ylab = "", main = "Color Palette Legend")

  for (i in 1:n_colors) {
    rect(0.2, legend_height - i + 0.25, 0.8, legend_height - i + 0.75,
         col = year_color_mapping[i], border = NA)
    text(1, legend_height - i + 0.5, names(year_color_mapping[i]), pos = 4)
  }
}
```

```
# Call the legend_plot function with the year_color_mapping
par(mfrow = c(1, 1))
par(mar = c(4, 4, 2, 2))
legend_plot(year_color_mapping)

### Citations ###
citation()
version$version.string
citation("ggplot2")
citation("dplyr")
citation("pROC")
citation('moments')
citation("jtools")
```

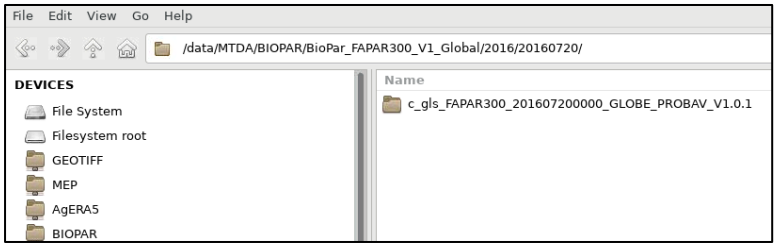
Addendum A

During defence of the thesis, it was raised by one of the examiners that since GDMP was derived from *f*APAR, it did not make sense that GDMP data was available prior to *f*APAR (Figure 5). The examiner suggested that the cause of this be sought and reported.

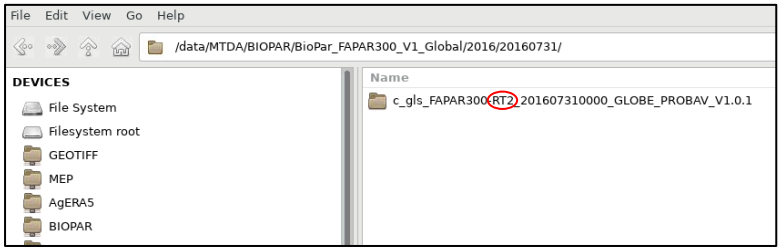
Subsequent investigation into this, revealed that the file naming convention used on the VITO NV remote VM changed in July 2016. Prior to July 2016, the file names did not contain the RT flag. Since the data extract bash script that I used in the study, derived data based on filenames, specifically RT2, and RT6; files before July 2016 were not extracted.

In summary, the *f*APAR data did exist, and corresponded with GDMP, it was just not selected due to an oversight in the way my script was written. Although the study may have benefited from having this additional data, by extending further in historical years, this oversight is not expected to affect the results significantly, because all "NA's" (missing data) were ignored during Correlation and Linear Regression, and the missing data is predominantly from PROBA-V, with little Sentinel-3 loss, caused by the transitional naming state. Reflecting on this, it may in fact have been prudent not to include PROBA-V data at all, eliminating another possible confounding factor of using different sensors for the start and end of the *f*APAR data.

Screenshot of VITO NV TerraScope Virtual Research Environment – 20/07/2016



Screenshot of VITO NV TerraScope Virtual Research Environment – 31/07/2016



Series from Lund University

Department of Physical Geography and Ecosystem Science

Master Thesis in Geographical Information Science

1. *Anthony Lawther:* The application of GIS-based binary logistic regression for slope failure susceptibility mapping in the Western Grampian Mountains, Scotland (2008).

2. *Rickard Hansen:* Daily mobility in Grenoble Metropolitan Region, France. Applied GIS methods in time geographical research (2008).

3. *Emil Bayramov:* Environmental monitoring of bio-restoration activities using GIS and Remote Sensing (2009).

4. *Rafael Villarreal Pacheco:* Applications of Geographic Information Systems as an analytical and visualization tool for mass real estate valuation: a case study of Fontibon District, Bogota, Columbia (2009).

5. *Siri Oestreich Waage:* a case study of route solving for oversized transport: The use of GIS functionalities in transport of transformers, as part of maintaining a reliable power infrastructure (2010).

6. *Edgar Pimiento:* Shallow landslide susceptibility – Modelling and validation (2010).

7. *Martina Schäfer:* Near real-time mapping of floodwater mosquito breeding sites using aerial photographs (2010).

8. *August Pieter van Waarden-Nagel:* Land use evaluation to assess the outcome of the programme of rehabilitation measures for the river Rhine in the Netherlands (2010).

9. *Samira Muhammad:* Development and implementation of air quality data mart for Ontario, Canada: A case study of air quality in Ontario using OLAP tool. (2010).

10. *Fredros Oketch Okumu*: Using remotely sensed data to explore spatial and temporal relationships between photosynthetic productivity of vegetation and malaria transmission intensities in selected parts of Africa (2011).

11. *Svajunas Plunge:* Advanced decision support methods for solving diffuse water pollution problems (2011).

12. *Jonathan Higgins:* Monitoring urban growth in greater Lagos: A case study using GIS to monitor the urban growth of Lagos 1990 - 2008 and produce future growth prospects for the city (2011).

13. *Mårten Karlberg:* Mobile Map Client API: Design and Implementation for Android (2011).

14. *Jeanette McBride:* Mapping Chicago area urban tree canopy using color infrared imagery (2011).

15. *Andrew Farina:* Exploring the relationship between land surface temperature and vegetation abundance for urban heat island mitigation in Seville, Spain (2011).

16. *David Kanyari*: Nairobi City Journey Planner: An online and a Mobile Application (2011).

17. *Laura V. Drews:* Multi-criteria GIS analysis for siting of small wind power plants - A case study from Berlin (2012).

18. *Qaisar Nadeem:* Best living neighborhood in the city - A GIS based multi criteria evaluation of ArRiyadh City (2012).

19. *Ahmed Mohamed El Saeid Mustafa:* Development of a photo voltaic building rooftop integration analysis tool for GIS for Dokki District, Cairo, Egypt (2012).

20. *Daniel Patrick Taylor*: Eastern Oyster Aquaculture: Estuarine Remediation via Site Suitability and Spatially Explicit Carrying Capacity Modeling in Virginia's Chesapeake Bay (2013).

21. *Angeleta Oveta Wilson:* A Participatory GIS approach to *unearthing* Manchester's Cultural Heritage '*gold mine'* (2013).

22. *Ola Svensson:* Visibility and Tholos Tombs in the Messenian Landscape: A Comparative Case Study of the Pylian Hinterlands and the Soulima Valley (2013).

23. *Monika Ogden:* Land use impact on water quality in two river systems in South Africa (2013).

24. *Stefan Rova:* A GIS based approach assessing phosphorus load impact on Lake Flaten in Salem, Sweden (2013).

25. *Yann Buhot:* Analysis of the history of landscape changes over a period of 200 years. How can we predict past landscape pattern scenario and the impact on habitat diversity? (2013).

26. *Christina Fotiou:* Evaluating habitat suitability and spectral heterogeneity models to predict weed species presence (2014).

27. *Inese Linuza:* Accuracy Assessment in Glacier Change Analysis (2014).

28. *Agnieszka Griffin:* Domestic energy consumption and social living standards: a GIS analysis within the Greater London Authority area (2014).

29. *Brynja Guðmundsdóttir:* Detection of potential arable land with remote sensing and GIS - A Case Study for Kjósarhreppur (2014).

30. *Oleksandr Nekrasov:* Processing of MODIS Vegetation Indices for analysis of agricultural droughts in the southern Ukraine between the years 2000-2012 (2014).

31. *Sarah Tressel:* Recommendations for a polar Earth science portal in the context of Arctic Spatial Data Infrastructure (2014).

32. *Caroline Gevaert:* Combining Hyperspectral UAV and Multispectral Formosat-2 Imagery for Precision Agriculture Applications (2014).

33. *Salem Jamal-Uddeen:* Using GeoTools to implement the multi-criteria evaluation analysis - weighted linear combination model (2014).

34. *Samanah Seyedi-Shandiz:* Schematic representation of geographical railway network at the Swedish Transport Administration (2014).

35. *Kazi Masel Ullah:* Urban Land-use planning using Geographical Information System and analytical hierarchy process: case study Dhaka City (2014).

36. *Alexia Chang-Wailing Spitteler:* Development of a web application based on MCDA and GIS for the decision support of river and floodplain rehabilitation projects (2014).

37. *Alessandro De Martino:* Geographic accessibility analysis and evaluation of potential changes to the public transportation system in the City of Milan (2014).

38. *Alireza Mollasalehi:* GIS Based Modelling for Fuel Reduction Using Controlled Burn in Australia. Case Study: Logan City, QLD (2015).

39. *Negin A. Sanati:* Chronic Kidney Disease Mortality in Costa Rica; Geographical Distribution, Spatial Analysis and Non-traditional Risk Factors (2015).

40. *Karen McIntyre:* Benthic mapping of the Bluefields Bay fish sanctuary, Jamaica (2015).

41. *Kees van Duijvendijk:* Feasibility of a low-cost weather sensor network for agricultural purposes: A preliminary assessment (2015).

42. *Sebastian Andersson Hylander:* Evaluation of cultural ecosystem services using GIS (2015).

43. *Deborah Bowyer:* Measuring Urban Growth, Urban Form and Accessibility as Indicators of Urban Sprawl in Hamilton, New Zealand (2015).

44. *Stefan Arvidsson:* Relationship between tree species composition and phenology extracted from satellite data in Swedish forests (2015).

45. *Damián Giménez Cruz*: GIS-based optimal localisation of beekeeping in rural Kenya (2016).

46. *Alejandra Narváez Vallejo:* Can the introduction of the topographic indices in LPJ-GUESS improve the spatial representation of environmental variables? (2016).

47. *Anna Lundgren:* Development of a method for mapping the highest coastline in Sweden using breaklines extracted from high resolution digital elevation models (2016).

48. *Oluwatomi Esther Adejoro:* Does location also matter? A spatial analysis of social achievements of young South Australians (2016).

49. *Hristo Dobrev Tomov:* Automated temporal NDVI analysis over the Middle East for the period 1982 - 2010 (2016).

50. *Vincent Muller:* Impact of Security Context on Mobile Clinic Activities A GIS Multi Criteria Evaluation based on an MSF Humanitarian Mission in Cameroon (2016).

51. *Gezahagn Negash Seboka:* Spatial Assessment of NDVI as an Indicator of Desertification in Ethiopia using Remote Sensing and GIS (2016).

52. *Holly Buhler:* Evaluation of Interfacility Medical Transport Journey Times in Southeastern British Columbia. (2016).

53. *Lars Ole Grottenberg*: Assessing the ability to share spatial data between emergency management organisations in the High North (2016).

54. *Sean Grant:* The Right Tree in the Right Place: Using GIS to Maximize the Net Benefits from Urban Forests (2016).

55. *Irshad Jamal:* Multi-Criteria GIS Analysis for School Site Selection in Gorno-Badakhshan Autonomous Oblast, Tajikistan (2016).

56. *Fulgencio Sanmartín:* Wisdom-volkano: A novel tool based on open GIS and time-series visualization to analyse and share volcanic data (2016).

57. *Nezha Acil:* Remote sensing-based monitoring of snow cover dynamics and its influence on vegetation growth in the Middle Atlas Mountains (2016).

58. *Julia Hjalmarsson:* A Weighty Issue: Estimation of Fire Size with Geographically Weighted Logistic Regression (2016).

59. *Mathewos Tamiru Amato:* Using multi-criteria evaluation and GIS for chronic food and nutrition insecurity indicators analysis in Ethiopia (2016).

60. *Karim Alaa El Din Mohamed Soliman El Attar:* Bicycling Suitability in Downtown, Cairo, Egypt (2016).

61. *Gilbert Akol Echelai:* Asset Management: Integrating GIS as a Decision Support Tool in Meter Management in National Water and Sewerage Corporation (2016).

62. *Terje Slinning:* Analytic comparison of multibeam echo soundings (2016).

63. *Gréta Hlín Sveinsdóttir:* GIS-based MCDA for decision support: A framework for wind farm siting in Iceland (2017).

64. *Jonas Sjögren:* Consequences of a flood in Kristianstad, Sweden: A GIS-based analysis of impacts on important societal functions (2017).

65. *Nadine Raska:* 3D geologic subsurface modelling within the Mackenzie Plain, Northwest Territories, Canada (2017).

66. *Panagiotis Symeonidis*: Study of spatial and temporal variation of atmospheric optical parameters and their relation with PM 2.5 concentration over Europe using GIS technologies (2017).

67. *Michaela Bobeck:* A GIS-based Multi-Criteria Decision Analysis of Wind Farm Site Suitability in New South Wales, Australia, from a Sustainable Development Perspective (2017).

68. *Raghdaa Eissa*: Developing a GIS Model for the Assessment of Outdoor Recreational Facilities in New Cities Case Study: Tenth of Ramadan City, Egypt (2017).

69. *Zahra Khais Shahid*: Biofuel plantations and isoprene emissions in Svea and Götaland (2017).

70. *Mirza Amir Liaquat Baig*: Using geographical information systems in epidemiology: Mapping and analyzing occurrence of diarrhea in urban - residential area of Islamabad, Pakistan (2017).

71. *Joakim Jörwall*: Quantitative model of Present and Future well-being in the EU-28: A spatial Multi-Criteria Evaluation of socioeconomic and climatic comfort factors (2017).

72. *Elin Haettner*: Energy Poverty in the Dublin Region: Modelling Geographies of Risk (2017).

73. *Harry Eriksson*: Geochemistry of stream plants and its statistical relations to soil- and bedrock geology, slope directions and till geochemistry. A GIS-analysis of small catchments in northern Sweden (2017).

74. *Daniel Gardevärn:* PPGIS and Public meetings – An evaluation of public participation methods for urban planning (2017).

75. *Kim Friberg:* Sensitivity Analysis and Calibration of Multi Energy Balance Land Surface Model Parameters (2017).

76. *Viktor Svanerud:* Taking the bus to the park? A study of accessibility to green areas in Gothenburg through different modes of transport (2017).

77. *Lisa-Gaye Greene*: Deadly Designs: The Impact of Road Design on Road Crash Patterns along Jamaica's North Coast Highway (2017).

78. *Katarina Jemec Parker*: Spatial and temporal analysis of fecal indicator bacteria concentrations in beach water in San Diego, California (2017).

79. *Angela Kabiru*: An Exploratory Study of Middle Stone Age and Later Stone Age Site Locations in Kenya's Central Rift Valley Using Landscape Analysis: A GIS Approach (2017).

80. *Kristean Björkmann*: Subjective Well-Being and Environment: A GIS-Based Analysis (2018).

81. *Williams Erhunmonmen Ojo*: Measuring spatial accessibility to healthcare for people living with HIV-AIDS in southern Nigeria (2018).

82. *Daniel Assefa*: Developing Data Extraction and Dynamic Data Visualization (Styling) Modules for Web GIS Risk Assessment System (WGRAS). (2018).

83. *Adela Nistora*: Inundation scenarios in a changing climate: assessing potential impacts of sea-level rise on the coast of South-East England (2018).

84. *Marc Seliger*: Thirsty landscapes - Investigating growing irrigation water consumption and potential conservation measures within Utah's largest master-planned community: Daybreak (2018).

85. *Luka Jovičić*: Spatial Data Harmonisation in Regional Context in Accordance with INSPIRE Implementing Rules (2018).

86. *Christina Kourdounouli*: Analysis of Urban Ecosystem Condition Indicators for the Large Urban Zones and City Cores in EU (2018).

87. *Jeremy Azzopardi*: Effect of distance measures and feature representations on distance-based accessibility measures (2018).

88. *Patrick Kabatha*: An open source web GIS tool for analysis and visualization of elephant GPS telemetry data, alongside environmental and anthropogenic variables (2018).

89. *Richard Alphonce Giliba*: Effects of Climate Change on Potential Geographical Distribution of Prunus africana (African cherry) in the Eastern Arc Mountain Forests of Tanzania (2018).

90. *Eiður Kristinn Eiðsson*: Transformation and linking of authoritative multi-scale geodata for the Semantic Web: A case study of Swedish national building data sets (2018).

91. *Niamh Harty*: HOP!: a PGIS and citizen science approach to monitoring the condition of upland paths (2018).

92. *José Estuardo Jara Alvear*: Solar photovoltaic potential to complement hydropower in Ecuador: A GIS-based framework of analysis (2018).

93. *Brendan O'Neill*: Multicriteria Site Suitability for Algal Biofuel Production Facilities (2018).

94. *Roman Spataru*: Spatial-temporal GIS analysis in public health – a case study of polio disease (2018).

95. *Alicja Miodońska*: Assessing evolution of ice caps in Suðurland, Iceland, in years 1986 - 2014, using multispectral satellite imagery (2019).

96. *Dennis Lindell Schettini*: A Spatial Analysis of Homicide Crime's Distribution and Association with Deprivation in Stockholm Between 2010-2017 (2019).

97. *Damiano Vesentini*: The Po Delta Biosphere Reserve: Management challenges and priorities deriving from anthropogenic pressure and sea level rise (2019).

98. *Emilie Arnesten*: Impacts of future sea level rise and high water on roads, railways and environmental objects: a GIS analysis of the potential effects of increasing sea levels and highest projected high water in Scania, Sweden (2019).

99. *Syed Muhammad Amir Raza*: Comparison of geospatial support in RDF stores: Evaluation for ICOS Carbon Portal metadata (2019).

100. *Hemin Tofiq*: Investigating the accuracy of Digital Elevation Models from UAV images in areas with low contrast: A sandy beach as a case study (2019).

101. *Evangelos Vafeiadis*: Exploring the distribution of accessibility by public transport using spatial analysis. A case study for retail concentrations and public hospitals in Athens (2019).

102. *Milan Sekulic*: Multi-Criteria GIS modelling for optimal alignment of roadway by-passes in the Tlokweng Planning Area, Botswana (2019).

103. *Ingrid Piirisaar*: A multi-criteria GIS analysis for siting of utility-scale photovoltaic solar plants in county Kilkenny, Ireland (2019).

104. *Nigel Fox*: Plant phenology and climate change: possible effect on the onset of various wild plant species' first flowering day in the UK (2019).

105. *Gunnar Hesch*: Linking conflict events and cropland development in Afghanistan, 2001 to 2011, using MODIS land cover data and Uppsala Conflict Data Programme (2019).

106. *Elijah Njoku*: Analysis of spatial-temporal pattern of Land Surface Temperature (LST) due to NDVI and elevation in Ilorin, Nigeria (2019).

107. *Katalin Bunyevácz*: Development of a GIS methodology to evaluate informal urban green areas for inclusion in a community governance program (2019).

108. *Paul dos Santos*: Automating synthetic trip data generation for an agent-based simulation of urban mobility (2019).

109. *Robert O' Dwyer*: Land cover changes in Southern Sweden from the mid-Holocene to present day: Insights for ecosystem service assessments (2019).

110. *Daniel Klingmyr*: Global scale patterns and trends in tropospheric NO2 concentrations (2019).

111. *Marwa Farouk Elkabbany*: Sea Level Rise Vulnerability Assessment for Abu Dhabi, United Arab Emirates (2019).

112. *Jip Jan van Zoonen*: Aspects of Error Quantification and Evaluation in Digital Elevation Models for Glacier Surfaces (2020).

113. *Georgios Efthymiou*: The use of bicycles in a mid-sized city – benefits and obstacles identified using a questionnaire and GIS (2020).

114. *Haruna Olayiwola Jimoh*: Assessment of Urban Sprawl in MOWE/IBAFO Axis of Ogun State using GIS Capabilities (2020).

115. *Nikolaos Barmpas Zachariadis*: Development of an iOS, Augmented Reality for disaster management (2020).

116. *Ida Storm*: ICOS Atmospheric Stations: Spatial Characterization of CO2 Footprint Areas and Evaluating the Uncertainties of Modelled CO2 Concentrations (2020).

117. *Alon Zuta*: Evaluation of water stress mapping methods in vineyards using airborne thermal imaging (2020).

118. *Marcus Eriksson*: Evaluating structural landscape development in the municipality Upplands-Bro, using landscape metrics indices (2020).

119. *Ane Rahbek Vierø*: Connectivity for Cyclists? A Network Analysis of Copenhagen's Bike Lanes (2020).

120. *Cecilia Baggini*: Changes in habitat suitability for three declining Anatidae species in saltmarshes on the Mersey estuary, North-West England (2020).

121. *Bakrad Balabanian*: Transportation and Its Effect on Student Performance (2020).

122. *Ali Al Farid*: Knowledge and Data Driven Approaches for Hydrocarbon Microseepage Characterizations: An Application of Satellite Remote Sensing (2020).

123. *Bartlomiej Kolodziejczyk*: Distribution Modelling of Gene Drive-Modified Mosquitoes and Their Effects on Wild Populations (2020).

124. *Alexis Cazorla*: Decreasing organic nitrogen concentrations in European water bodies - links to organic carbon trends and land cover (2020).

125. *Kharid Mwakoba*: Remote sensing analysis of land cover/use conditions of community-based wildlife conservation areas in Tanzania (2021).

126. *Chinatsu Endo*: Remote Sensing Based Pre-Season Yellow Rust Early Warning in Oromia, Ethiopia (2021).

127. *Berit Mohr*: Using remote sensing and land abandonment as a proxy for long-term human out-migration. A Case Study: Al-Hassakeh Governorate, Syria (2021).

128. *Kanchana Nirmali Bandaranayake*: Considering future precipitation in delineation locations for water storage systems - Case study Sri Lanka (2021).

129. *Emma Bylund*: Dynamics of net primary production and food availability in the aftermath of the 2004 and 2007 desert locust outbreaks in Niger and Yemen (2021).

130. *Shawn Pace*: Urban infrastructure inundation risk from permanent sea-level rise scenarios in London (UK), Bangkok (Thailand) and Mumbai (India): A comparative analysis (2021).

131. *Oskar Evert Johansson*: The hydrodynamic impacts of Estuarine Oyster reefs, and the application of drone technology to this study (2021).

132. *Pritam Kumarsingh*: A Case Study to develop and test GIS/SDSS methods to assess the production capacity of a Cocoa Site in Trinidad and Tobago (2021).

133. *Muhammad Imran Khan*: Property Tax Mapping and Assessment using GIS (2021).

134. *Domna Kanari*: Mining geosocial data from Flickr to explore tourism patterns: The case study of Athens (2021).

135. *Mona Tykesson Klubien*: Livestock-MRSA in Danish pig farms (2021).

136. *Ove Njøten*: Comparing radar satellites. Use of Sentinel-1 leads to an increase in oil spill alerts in Norwegian waters (2021).

137. *Panagiotis Patrinos*: Change of heating fuel consumption patterns produced by the economic crisis in Greece (2021).

138. *Lukasz Langowski*: Assessing the suitability of using Sentinel-1A SAR multi-temporal imagery to detect fallow periods between rice crops (2021).

139. *Jonas Tillman*: Perception accuracy and user acceptance of legend designs for opacity data mapping in GIS (2022).

140. *Gabriela Olekszyk*: ALS (Airborne LIDAR) accuracy: Can potential low data quality of ground points be modelled/detected? Case study of 2016 LIDAR capture over Auckland, New Zealand (2022).

141. *Luke Aspland*: Weights of Evidence Predictive Modelling in Archaeology (2022).

142. *Luís Fareleira Gomes*: The influence of climate, population density, tree species and land cover on fire pattern in mainland Portugal (2022).

143. *Andreas Eriksson*: Mapping Fire Salamander (Salamandra salamandra) Habitat Suitability in Baden-Württemberg with Multi-Temporal Sentinel-1 and Sentinel-2 Imagery (2022).

144. *Lisbet Hougaard Baklid*: Geographical expansion rate of a brown bear population in Fennoscandia and the factors explaining the directional variations (2022).

145. *Victoria Persson*: Mussels in deep water with climate change: Spatial distribution of mussel (Mytilus galloprovincialis) growth offshore in the French Mediterranean with respect to climate change scenario RCP 8.5 Long Term and Integrated Multi-Trophic Aquaculture (IMTA) using Dynamic Energy Budget (DEB) modelling (2022).

146. *Benjamin Bernard Fabien Gérard Borgeais*: Implementing a multi-criteria GIS analysis and predictive modelling to locate Upper Palaeolithic decorated caves in the Périgord noir, France (2022).

147. *Bernat Dorado-Guerrero*: Assessing the impact of post-fire restoration interventions using spectral vegetation indices: A case study in El Bruc, Spain (2022).

148. *Ignatius Gabriel Aloysius Maria Perera*: The Influence of Natural Radon Occurrence on the Severity of the COVID-19 Pandemic in Germany: A Spatial Analysis (2022).

149. *Mark Overton*: An Analysis of Spatially-enabled Mobile Decision Support Systems in a Collaborative Decision-Making Environment (2022).

150. *Viggo Lunde*: Analysing methods for visualizing time-series datasets in open-source web mapping (2022).

151. *Johan Viscarra Hansson*: Distribution Analysis of Impatiens glandulifera in Kronoberg County and a Pest Risk Map for Alvesta Municipality (2022).

152. *Vincenzo Poppiti*: GIS and Tourism: Developing strategies for new touristic flows after the Covid-19 pandemic (2022).

153. *Henrik Hagelin*: Wildfire growth modelling in Sweden - A suitability assessment of available data (2023).

154. *Gabriel Romeo Ferriols Pavico*: Where there is road, there is fire (influence): An exploratory study on the influence of roads in the spatial patterns of Swedish wildfires of 2018 (2023).

155. *Colin Robert Potter*: Using a GIS to enable an economic, land use and energy output comparison between small wind powered turbines and large-scale wind farms: the case of Oslo, Norway (2023).

156. *Krystyna Muszel*: Impact of Sea Surface Temperature and Salinity on Phytoplankton blooms phenology in the North Sea (2023).

157. *Tobias Rydlinge*: Urban tree canopy mapping - an open source deep learning approach (2023).

158. *Albert Wellendorf*: Multi-scale Bark Beetle Predictions Using Machine Learning (2023).

159. *Manolis Papadakis*: Use of Satellite Remote Sensing for Detecting Archaeological Features: An Example from Ancient Corinth, Greece (2023).

160. *Konstantinos Sourlamtas*: Developing a Geographical Information System for a water and sewer network, for monitoring, identification and leak repair - Case study: Municipal Water Company of Naoussa, Greece (2023).

161. *Xiaoming Wang*: Identification of restoration hotspots in landscape-scale green infrastructure planning based on model-predicted connectivity forest (2023).

162. *Sarah Sienaert*: Usability of Sentinel-1 C-band VV and VH SAR data for the detection of flooded oil palm (2023).

163. *Katarina Ekeroot*: Uncovering the spatial relationships between Covid-19 vaccine coverage and local politics in Sweden (2023).

164. *Nikolaos Kouskoulis*: Exploring patterns in risk factors for bark beetle attack during outbreaks triggered by drought stress with harvester data on attacked trees: A case study in Southeastern Sweden (2023).

165. *Jonas Almén*: Geographic polarization and clustering of partisan voting: A local-level analysis of Stockholm Municipality (2023).

166. *Sara Sharon Jones*: Tree species impact on Forest Fire Spread Susceptibility in Sweden (2023).

167. *Takura Matswetu*: Towards a Geographic Information Systems and Data-Driven Integration Management. Studying holistic integration through spatial accessibility of services in Tampere, Finland. (2023).

168. *Duncan Jones*: Investigating the influence of the tidal regime on harbour porpoise *Phocoena phocoena* distribution in Mount's Bay, Cornwall (2023).

169. *Jason Craig Joubert*: A comparison of remote sensed semi-arid grassland vegetation anomalies detected using MODIS and Sentinel-3, with anomalies in ground-based eddy covariance flux measurements (2023).