

Faculté des bioingénieurs

Improving maize crop yield estimation by assimilating remotely sensed biophysical variables in AquaCrop agrometeorological model: A case study of Western Kenya

Auteur: Catherine Akinyi Odera

Promoteur(s): Prof. Pierre Defourny (UCL/ELI/ELIE)
Prof. Joost Wellens (University of Liège)

Lecteur(s): Prof. Xavier Draye (UCL/ELI/ELIE)
Dr. Sophie Bontemps (UCL/ELI/ELIE)

Année académique 2022-2023

Mémoire de fin d'études présenté en vue de l'obtention du diplôme de

Master SAIV – MSc Erasmus+ GEM: Geo-information Science and Earth Observation for Environmental Modelling and Management

Table of Contents

Abstract	i
Acknowledgement	ii
List of Figures	iii
List of Tables	v
List of Equations	vi
List of Abbreviations	vii
CHAPTER 1. Introduction.....	1
CHAPTER 2. Literature Review	4
2.1 Food Security and Climate Change	4
2.2. Crop growth modelling and Satellite dataset.....	5
2.3. AquaCrop growth model	8
2.4. Assimilation of biophysical variables.....	12
CHAPTER 3. Research Objectives.....	15
CHAPTER 4. Materials and Methods	16
4.1. Study area	16
4.2. Methodology.....	17
4.3. Tools	19
4.4. Data and processing.....	19
4.4.1. Maize yield dataset.....	20
4.4.2. Parcel data verification.....	23
4.4.3. Satellite dataset.....	23
4.4.4. FCover processing.....	23
4.4.5. Cloud masking.....	24
4.4.6. FCover values extraction, gap filling and smoothing	24
4.4.7. Weather dataset	25
4.4.8. Weather data processing.....	26
4.4.9. Soil texture dataset	28
4.4.10. Soil data processing.....	28
4.4.11. Aquacrop data and yield simulation.....	29
4.4.12. Model sensitivity analysis	30

4.4.13. Assimilation of FCover in AquaCrop	30
4.4.14. Model performance evaluation.....	31
4.4.15. Data aggregation	33
CHAPTER 5. Results and Discussion	35
5.1. Results	35
5.1.1. Parcel yield verification results	35
5.1.2. Parcel yield analysis results.....	37
5.1.3. FCover values.....	38
5.1.4. Relationship between yield and FCover.....	41
5.1.5. Sensitivity analysis results	44
5.1.6. Cloud masking results	46
5.1.7. Yield simulation results before assimilation	46
5.1.8. Yield simulation results after assimilation	50
5.1.9. Yield simulation results at ward level	53
5.2. Discussion.....	56
5.2.1. Parcel heterogeneity and diversity of crop varieties	56
5.2.2. Model Sensitivity Analysis	56
5.2.3. FCover Time Series.....	57
CHAPTER 6. Conclusion and Recommendations.....	60
6.1. Conclusion	60
6.2. Recommendations	60
CHAPTER 7. References.....	A
CHAPTER 8. Appendix.....	G
Appendix 1: Parcels' FCover profiles	G
Appendix 2: Elevation Map of Keiyo Marakwet and Uasin Gishu.....	JJ
Appendix 3: Uasin Gishu and Keiyo Marakwet landforms	KK
Appendix 4: Soil classification.....	LL
Appendix 5: Model performance before assimilation according to planting dates	LL
Appendix 6: Model performance after assimilation according to planting dates.....	NN

Abstract

Crop yield estimation is vital in providing insights to policymakers on the state of food security within a given country. The application of crop yield estimation modelling has been realised to provide important information for data-driven decision-making. However, the models used have some shortcomings in providing accurate yield estimations. Studies have shown that yield estimation significantly improves by assimilating remotely sensed biophysical variables into the crop estimation models to curb this challenge. This study employs AquaCrop agro-meteorological model in maize crop yield estimation, following its successful application in previous studies in arid and semi-arid areas (ASALs). The study addresses the simulation of maize yield under rain-fed conditions in Western Kenya within two counties, i.e., Uasin Gishu and Keiyo Marakwet, for 2021. The meteorological data used in the model was obtained from ERA5-land, which included precipitation, maximum temperature, minimum temperature, and potential evapotranspiration. In addition, satellite-derived FCover obtained from Sentinel-2 was assimilated into the model to improve its performance. The model simulation performance of maize yield before FCover assimilation was $R^2 = 0.198$ at the parcel level and $R^2 = 0.812$ at the ward level. After the FCover assimilation, the accuracy improved to $R^2 = 0.334$ at the parcel level and $R^2 = 0.843$ at the ward level. These results demonstrate that remotely sensed biophysical variables improve the performance of crop yield estimation models.

KEYWORDS: Kenya, Food security, Maize, AquaCrop, Yield Estimation, Sentinel-2, FCover

Acknowledgement

First and foremost, I am profoundly grateful to God for being my constant source of strength, inspiration, wisdom, and knowledge throughout my academic journey. His unwavering presence has guided me every step of the way.

I extend my heartfelt appreciation to my supervisors, Professor Pierre Defourny, Professor Joost Wellens and Pierre Hudmont, for their unwavering support, constructive criticism, motivation, and guidance during this project. Their expertise played an invaluable role in shaping the trajectory of this study and enhancing my skills in the process. I am grateful for the time and effort they dedicated.

I would like to express my gratitude to my jury members, Professor Xavier Draye and Doctor Sophie Bontemps, for their insightful comments and constructive criticism that have helped refine my ideas and strengthen my arguments. Their expertise and willingness to share their knowledge in this research area have been instrumental in my academic growth.

I would like to acknowledge the GEM Erasmus Mundus Program in partnership with the University of Twente, the University of Tartu, and the University Catholique de Louvain for providing me with the resources and opportunity to pursue my academic goals. I am grateful for the knowledge and skills I have gained during this academic journey.

My gratitude goes to the Bird-Eye project's partners (PULA, Tetra Tech and the Bill & Melinda Gates Foundation) for their open data policy that made this study possible. In addition, I would like to thank Mr Steve Kory Korir (Agriculture data scientist) at PULA, Kenya, for his assistance and unwavering guidance on small-scale agricultural practices in Kenya.

Finally, I would like to express my heartfelt gratitude to my family, classmates and friends for their unwavering support and encouragement throughout this journey. Your faith in me and belief in my abilities have motivated and inspired me.

List of Figures

Figure 1: Physiological processes in the AquaCrop model (Vanuytrecht et al., 2014)	9
Figure 2: Variation of green canopy cover throughout the growing cycle under non-stress conditions (FAO, 2022)	10
Figure 3: Schematic illustration of the different methods for the assimilation of remotely sensed biophysical variables in crop yield estimation models (Delécolle et al., 1992)	12
Figure 4: Study area map showing the location of Uasin Gishu and Keiyo Marakwet.....	17
Figure 5: Methodology schema	18
Figure 6: Box placement exercise within a parcel (Bird's-Eye).....	20
Figure 7: Drying and weighing of the fresh harvest (Bird's-Eye)	21
Figure 8: Yield distribution with respect to elevation and soil texture	22
Figure 9: Features detected by FMask	24
Figure 10: Temperature data profile for parcel code_0 in 2021	25
Figure 11: Precipitation data profile for parcel code_0 in 2021	26
Figure 12: Potential evapotranspiration data profile for parcel code_0 in 2021.....	26
Figure 13: Soil texture triangle defined by USDA	28
Figure 14: Map showing the distribution of soil textures in Uasin Gishu and Keiyo Marakwet .	29
Figure 15: Uasin Gishu and Keiyo Marakwet Wards considered for yield aggregation	34
Figure 16: Distribution of maize yield for selected parcels in 2021	35
Figure 17: Parcels' heterogeneity	36
Figure 18: FCover and True colour composite of parcel code_73 showing different vegetation density	36
Figure 19: Crop cycle duration against planting dates.....	37
Figure 20: Maximum FCover by soil texture	38
Figure 21: Relationship between elevation and crop cycle duration	38
Figure 22: Parcels' FCover profiles with the orange line representing the maximum FCover values	39
Figure 23: Parcel with an undesirable FCover time series	40
Figure 24: FCover profiles with the red line showing the emergence date	40
Figure 25: Relationship between the reported yield and maximum FCover for 108 parcels	42
Figure 26: Relationship between the reported yield and maximum FCover for parcels with planting dates between March 3-26	43
Figure 27: Relationship between the reported yield and maximum FCover for parcels with planting dates between March 27- April 14.....	43
Figure 28: Relationship between the reported yield and maximum FCover for parcels with planting dates between April 15-27, 2021	44
Figure 29: Yield simulation with Clay soil compared with the reported yield.....	45
Figure 30: Yield simulation with Sandy clay loam soil compared with reported yield	45

Figure 31: FCover image before (left) and after (right) cloud masking	46
Figure 32: AquaCrop model yield estimation performance before the assimilation of FCover...	47
Figure 33: AquaCrop model yield estimation performance before the assimilation for March 3-26 planting dates	48
Figure 34: Simulated and reported yield against crop cycle duration	49
Figure 35: Simulated and reported yield against planting dates	49
Figure 36: Model yield estimation performance after the assimilation of FCover	51
Figure 37: Model yield estimation performance after assimilation for March 3-26 planting dates	51
Figure 38: Simulated (assimilation) and reported yield against crop cycle duration	52
Figure 39: Relationship between simulated yield after assimilation and Maximum FCover	52
Figure 40: Model yield performance before assimilation aggregated at the ward level.....	53
Figure 41: Model yield performance after assimilation aggregated at the ward level.....	54

List of Tables

Table 1: Percentage change in cereal production in Kenya (2020-2021)(FAO, 2023)	2
Table 2: Dataset summary	19
Table 3: Parcels' distribution within the planting and harvesting months.....	22
Table 4: AquaCrop parameters	30
Table 5: Test dataset containing Maximum FCover and duration between crop emergence and maximum FCover	41
Table 6: Yield estimated at the ward level after aggregation	54
Table 7: Summary of the model performance	55
Table 8: Summary of the comparison of correlation coefficient using the Fisher exact test.....	55

List of Equations

Equation 1: Growing degree days equation	10
Equation 2: Crop transpiration equation	11
Equation 3: Above-ground biomass equation.....	11
Equation 4: Yield calculation equation.....	11
Equation 5: Standard deviation equation	23
Equation 6: Penman-Monteith equation	27
Equation 7: Emergence FCover value calculation.....	31
Equation 8: MSE equation	31
Equation 9: MAE equation	32
Equation 10: RMSE equation	32
Equation 11: R^2 equation	32
Equation 12: NSE equation.....	33
Equation 13: Correlation coefficient equation.....	33
Equation 14: Test statistics equation.....	33
Equation 15: Yield aggregation calculation equation	34

List of Abbreviations

AEZ	Agroecological Zones
AGB	Above Ground Biomass
ANN	Artificial Neural Network
ASAL	Arid and Semi-Arid Lands
CC	Canopy Cover
CCAFS	Climate Change Agriculture and Food Security
CDC	Canopy Decline Coefficient
CGC	Canopy Growth Coefficient
CGSMs	Crop Growth Simulation Models
CropSyst	Cropping Systems Simulation Model
DEM	Digital Elevation Model
EnKF	Ensemble Kalman Filter
EKF	Extended Kalman Filter
ET _o	Potential evapotranspiration
FAO	Food and Agriculture Organization
FCover	Fractional canopy cover
FEWS-NET	Famine Early Warning System
GIEWS	Global Information and Early Warning System on Food and Agriculture
GDD	Growing Degree Days
HI	Harvesting Index
HWSD	Harmonized World Soil Database
ICCYF	Canadian Crop Yield Forecaster
ICRISAT	International Crops Research Institute for the Semi-Arid Tropics
KENSOTER	Soil and Terrain Database for Kenya
KF	Kalman Filter
LAI	Leaf Area Index
MAE	Mean Actual Error
MSE	Mean Squared Error
NSE	Nash-Sutcliffe Efficiency
P.F	Particle Filter
PSO	Particle Swarm Optimization
RMSE	Root Mean Squared Error
SAFY	Simple Algorithm for Yield Estimation
SDGs	Sustainable Development Goals
U.N	United Nations
USDA	United States Department of Agriculture
WFP	World Food Programme
WOFOST	World Food Studies

CHAPTER 1. Introduction

Maize is Kenya's staple food and is classified as a cash crop and a food crop (Kipkorir et al., 2007). It is a major source of nutrition for millions of Kenyans, and there is a growing demand for increased maize production due to the population growth in the country (Midega et al., 2016). Increasing maize production to meet the population demand will ensure the country's food security. Maize produced in Kenya is consumed within the country, and some are exported, contributing to 26% of Kenya's Gross Domestic Product (GDP) (Nakalembe et al., 2021).

Farming of maize in Kenya is predominantly practised in the western part of the country within the Rift Valley region, also referred to as the “grain basket” of Kenya (Kipkorir et al., 2007). Agriculture in this region relies heavily on rainfall, contributing to 75% of the total maize produced in Kenya (Okoth et al., 2012). In Uasin Gishu and Elgeyo Marakwet, farmers engage in small- and large-scale farming based on land availability. The commonly cultivated maize varieties are hybrids, particularly Hybrid-614 (H614), with a 41.53% adoption rate among farmers, followed by H6213 at 39.87% (Gachara et al., 2022). Some farmers also choose Indigenous maize varieties like Ndume, Pannar, and Duma due to their disease resistance, high productivity levels and larger cob sizes (Gachara et al., 2022).

The success of agriculture in Kenya's Rift Valley region is directly attributed to the favourable climatic conditions, making it an ideal location for maize cultivation. However, climate change is the main culprit that constrains agricultural production, resulting in unpredictable rainfall seasons and severe and persistent drought (Omoyo et al., 2015). The impact of climate change has led to decreased yield output and disruption of the planning of farming sowing dates (Mugalavai et al., 2008). Additionally, the landscape is characterized by varying topography and soil types (Moore et al., 2012) within the Kenyan agroecological zones (AEZ). The variation in topography significantly affects the crop yield outcome obtained for a given parcel (Moore et al., 2012).

Accurate estimation of maize yield is therefore crucial for agricultural decision-making, including crop management practices and policy formulation. Traditional methods of crop yield estimation that are typically used are often time-consuming, expensive, and require extensive field data

collection (Lobell et al., 2020; Mohamed Sallah et al., 2019; Karimli and Selbesoğlu, 2023). In a bid to address the challenges of crop yield estimation using traditional methods, remote sensing has emerged as a promising technology for crop yield estimation, especially in heterogeneous landscapes (Chivasa et al., 2017; Ozdogan et al., 2010). Satellite images provide valuable information on crop biophysical variables such as Fractional Canopy Cover (FCover) and Leaf Area Index (LAI), which can be coupled with crop yield estimation models to improve their performance. By incorporating remote sensing data into the analysis, researchers and policymakers can make more informed decisions regarding crop management and policy implementation.

A report produced by the Food and Agriculture Organization (FAO)/Global Information and Early Warning System (GIEWS) (see Table 1) shows a decline in maize production in Kenya from 2020 to 2021 by 14%, demonstrating the importance of obtaining crop yield estimation information. According to the World Food Programme (WFP) (2022), the decline in maize production in Kenya in 2021 was due to increased fertilizer prices and the drought that has affected the region since 2014, causing a decline in agriculture production. Furthermore, the increase in fertilizer prices and the impact of the drought were projected to lead to a decline in maize production in 2022 by an additional 12% compared to 2021 (World Food Programme, 2022).

Table 1: Percentage change in cereal production in Kenya (2020-2021)(FAO, 2023)

Kenya				
Cereal Production				
	2016-2020 average	2020	2021 estimate	change 2021/2020
	000 tonnes			percent
Maize	3 623	3 780	3 250	-14.0
Wheat	284	300	250	-16.7
Sorghum	222	250	240	-4.0
Others	308	361	314	-13.0
Total	4 437	4 691	4 054	-13.6

This decline in production emphasizes the need for improved crop yield estimation methods to effectively monitor and predict changes in maize yield for a particular growing season. By enhancing our understanding of the relationship between remotely sensed biophysical variables and crop yield, we can develop more reliable and efficient models for estimating maize crop yield.

This study aims at improving maize crop yield estimation in Western Kenya by assimilating remotely sensed biophysical variables in the AquaCrop agrometeorological model. AquaCrop is a widely used crop growth model that simulates crop yield based on various input parameters, including weather data, soil characteristics, and crop management practices (Steduto et al., 2009). Integrating remotely sensed data has been proven to enhance the model's accuracy by providing information on crop growth and development (Abi Saab et al., 2021).

The study presents a case of Western Kenya, where maize is grown on large extents of land. By focusing on this location, the study aims to capture the unique challenges and opportunities associated with maize production in this geographic region. The results of this study will provide valuable insights into the potential of coupling remote sensing and crop yield estimation models in improving yield estimation and enhancing food security in African countries.

CHAPTER 2. Literature Review

This chapter focuses on studies already undertaken in yield estimation, specifically concentrating on the AquaCrop growth model. The chapter touches on food security before discussing the role of crop growth models in yield estimation. Finally, the section focuses on the AquaCrop growth model and discusses the impact of the assimilation of remotely sensed biophysical variables in crop yield models' performance.

2.1 Food Security and Climate Change

FAO estimates that nearly 900 million people worldwide suffer from chronic hunger (Dawson et al., 2016). In 2015, the United Nations (U.N.) established the 17 Sustainable Development Goals (SDGs), with the essential goal of Zero Hunger (SDG 2) (Viana et al., 2022). The target of SDG 2 was to decrease the number of people suffering from hunger in the world by half. However, projections of rapid population growth coupled with global climate change have significantly negative impacts on food security (Dawson et al., 2016). In addition to growing population pressure in many parts of the world, food production needs to be increased through better water management and efficient food production systems that effectively meet the high population demand (Araya et al., 2010a; Mibulo and Kiggundu, 2018).

According to the European Commission Joint Research Centre Institute for Environment and Sustainability and CCAFS(Climate Change Agriculture and Food Security) (2012), many factors contribute to pressure on the global food supply. These factors range from underinvestment in the agricultural sector, climate change, climate variability, water scarcity, and drought to increased food demand and dietary changes. Feeding its growing population is a major challenge for the world, and it is easy to see that agriculture is the best answer to this challenge. Given global environmental problems, growing more food is not enough; therefore, agriculture must be managed sustainably (Lindblom et al., 2017; Welch and Graham, 1999).

Uncertainty about agricultural production and investment in agricultural improvements in African countries such as Kenya has raised concerns among local authorities and global bodies. Sufficient attention must be paid to rainfed agriculture as a critical element of food security (Kenya

Agricultural Research Laboratories-Kenya et al., 2007). Accurate crop yield measurement is required to understand the determinants of agricultural productivity. In smallholder production systems in third-world countries, yields have traditionally been acquired based on farmer-reported production or farm surveys, sometimes through obtaining crop cuttings for a subsection of an agricultural plot (Lobell et al., 2020), which may not be reliable at times.

2.2. Crop growth modelling and Satellite dataset

The traditional approach to estimating crop yields is generally reliable for regions like districts (Mitran et al., 2021). Nevertheless, when more precise crop statistics are needed for smaller areas, a significantly larger number of crop cuttings are normally required. As a result, this method becomes costly due to increased expenses (Mitran et al., 2021). Estimating crop production for extensive spatial scales like countries requires considerable effort, leading to high costs and time constraints. Therefore, utilising remote sensing and modelling-based methods for estimating crop yield provides a highly efficient and effective alternative solution (Mitran et al., 2021; Qader et al., 2021).

Crop growth simulation models (CGSMs) have advanced in their applications in recent decades due to the increased availability of data needed to simulate plant development and growth (Lobell et al., 2020). In addition, data collection technologies, e.g., satellites, have advanced almost exponentially, providing high-quality data on the state of ground vegetation at a particular time. As a result, satellite data continues to improve in terms of the spatial, temporal, and spectral resolution needed to detect performance on agricultural parcels (Lobell et al., 2020; Nakalembe et al., 2021). One such satellite instrument, Sentinel-2, has emerged to be useful in monitoring land surface changes due to its high spatial and spectral resolution (Szostak et al., 2020). Specifically, retrieving FCover values provides detailed information on vegetation density and distribution over large areas (Waldner et al., 2015).

CGSMs that have been employed in estimating crop yield in various studies include models such as AquaCrop (Steduto et al., 2009), WOFOST (World Food Studies) (Curnel et al., 2011), CropSyst (Cropping Systems Simulation Model) (Stöckle et al., 2003) and SAFY (Simple Algorithm for Yield Estimation) (Chahbi et al., 2014). Studies have used these models to simulate and predict crop yields based on input parameters such as climate data, soil texture, and crop

information. The AquaCrop model, for example, focuses on water-limited crop growth and yield estimation. It incorporates various parameters related to crop physiology, soil water dynamics, and the effect of water stress on yield formation (Steduto et al., 2009). WOFOST, on the other hand, is a well-established model that integrates weather data, soil properties, and crop-specific information to estimate crop growth, development, and yield (Curnel et al., 2011). CropSyst is a crop growth model that simulates various aspects of crop growth, water balance, nutrient cycling, and pest dynamics (Stöckle et al., 2003). SAFY is a simplified model that utilises a simple algorithm to estimate crop yield by considering parameters such as radiation, temperature, and soil moisture (Chahbi et al., 2014). The models are useful for researchers, agronomists, and policymakers to assess and monitor crop production in different regions of the world.

In addition to the models, several platforms and systems exist that utilise a combination of remotely sensed data, statistical models, and crop simulation models to provide information on crop conditions, food pricing, and yield forecasts (Basso and Liu, 2019) at both global and regional scales. These platforms are important in improving agricultural decision-making and enhancing food security. Some examples of such platforms include MARS (Monitoring Agricultural Resources) bulletin, developed by the European Commission's Joint Research Centre and combines satellite data, weather information, and crop models to provide early warning on crop yield and production in Europe (van der Velde et al., 2019). MARS offers timely and accurate information to support policy formulation and market analysis. Global Information and Early Warning System on Food and Agriculture (GIEWS), operated by the Food and Agriculture Organization (FAO), monitors global food production and provides early warnings on potential food security issues. It combines remote sensing data, field surveys, and statistical analysis to assess crop conditions, food supply, and market trends (Rashid, 2003).

GEOGLAM (Global Agricultural Monitoring) is another example of a platform that uses earth observation data and models to monitor agricultural production and provide reliable information on crop conditions and yield forecasts. GEOGLAM focuses on major staple crops and provides detailed assessments at regional and global levels (Becker-Reshef et al., 2019). CropWatch is another additional platform developed by the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT). It is a web-based platform that integrates satellite data, climate

information, and crop models to monitor and forecast crop performance in different regions. It supports decision-making for farmers, extension workers, and policymakers by providing information on crop growth, water requirements, and yield potential (Gommès et al., 2017). The Canadian Crop Yield Forecaster (ICCYF), developed by Agriculture and Agri-Food Canada, also uses satellite data, weather information, and crop models to generate timely and accurate crop yield forecasts for major Canadian crops. ICCYF helps in risk assessment, market analysis, and policy development (Chipanshi et al., 2015). Finally, Famine Early Warning System (FEWS-NET), funded by the United States Agency for International Development (USAID), combines remote sensing data, ground observations, and statistical models to provide early warnings and analysis of food security conditions in vulnerable regions. It helps decision-makers and humanitarian organizations to respond effectively to food crises (Becker-Reshef et al., 2010).

Crop growth monitoring has started to be effectively used in crop growth assessment, considering various environmental factors and management decisions (Mohamed Sallah et al., 2019). To improve the estimates, including valuable information from remote sensing measurements in the simulation processes of the crop model is of particular importance (Abi Saab et al., 2021). It is generally recognized that yield estimates obtained using crop yield estimation models can be improved by assimilating biophysical variables estimations, e.g., leaf area index, transpiration, FCover, soil moisture and surface temperature during the growing season at relevant spatial and temporal resolutions (de Wit et al., 2012; Weiss et al., 2020).

Using the biophysical variables of plant growth, crop simulation models can be utilised to understand the effects of weather conditions on plant growth during the growing season (Vanuytrecht et al., 2014). Satellite-based approaches to mapping agricultural outcomes, such as crop-specific yield and crop estimates, require data to train and validate the underlying remote sensing models.

Operational systems for regional crop monitoring and yield forecasting are generally based on an integrated analysis of meteorological data. Within such systems, the first level of analysis usually concerns weather conditions, as the climate is the primary driver of crop growth and an essential factor behind annual yield variability (Mandal and Rao, 2020). However, yield predictions based

solely on meteorological variables are generally only possible in regions where yield is determined by a single strong indicator, such as the total amount of precipitation (Vanuytrecht et al., 2014).

In third-world countries, agriculture is majorly undertaken on a small scale and is managed by households. Several aspects of small-scale agriculture, beyond small fields, pose a significant challenge in remote sensing (Jin et al., 2019). First, there is cloud cover in many small-scale agricultural areas, mainly in tropical regions where crops are actively grown during the rainy season. Clouds make it challenging to obtain a complete time series covering the stages of the growing cycle (Delrue et al., 2013). Second, most smallholder systems have enormous heterogeneity since families typically grow multiple crops on the same parcel. Furthermore, there are often significant differences between crops grown by smallholders in a region and farmers in neighbouring districts, counties, or states (Jin et al., 2019).

The different studies on crop yield simulation using CGSMs often produce different outcomes due to the complex nature of crop growth and yield and the number of factors that can affect them (Roell et al., 2020). These factors can include but are not limited to weather patterns, changes in soil nutrients, and pest and disease infestations. In addition, applying varied management practices and using advanced agricultural technology may also impact the model's performance (Roell et al., 2020). Furthermore, differences in model parameters, assumptions, and input data quality or quantity variations can contribute to different crop yield simulation results (Roell et al., 2020).

2.3. AquaCrop growth model

Water has always been the main factor limiting agricultural production in most regions with insufficient rainfall (Lobell et al., 2011; Latiri et al., 2010; Rockström et al., 2010). With the increase in agricultural demands, the need to improve efficient water use in agricultural production has increased (Hsiao et al., 2009). To achieve this goal, FAO in 2009 developed a crop simulation model called AquaCrop (Steduto et al., 2009). The model strikes a balance between accuracy, simplicity, robustness, and ease of use (Hsiao et al., 2009; Mibulo and Kiggundu, 2018).

AquaCrop is a water-driven crop simulation model that simulates yield responses to water (Foster et al., 2017). Operation of the AquaCrop model requires input data consisting of climate, crop, soil, field parameters and irrigation management information. The model contains a complete set

of input parameters that can be selected and adjusted for different soil types or crops (Abedinpour et al., 2012). The three main differences between AquaCrop and other crop yield estimation models are the emphasis on the use of crop water, the expression of FCover rather than the LAI, and the separation of evaporation in the soil and plant transpiration in the context of the soil water balance of the root zone and normalised water productivity (Steduto et al. 2009). During crop establishment, when land cover is still very low, soil evaporation can be a rather important part of evapotranspiration, especially in arid areas (Bello and Walker, 2016), which is taken into account by Aquacrop.

The AquaCrop model simulates the variation in accessible plant biomass and crop yield in response to variations in soil moisture in the root zone, as shown in Figure 1. Simulations are done in daily time steps considering incoming and outgoing water flows and taking into account the daily transpiration rate (Araya et al., 2010b; Steduto et al., 2009). To understand the projected future climate change effects in agricultural areas, model calibration could be crucial in generating yield predictions and improving water use (Araya et al., 2010a).

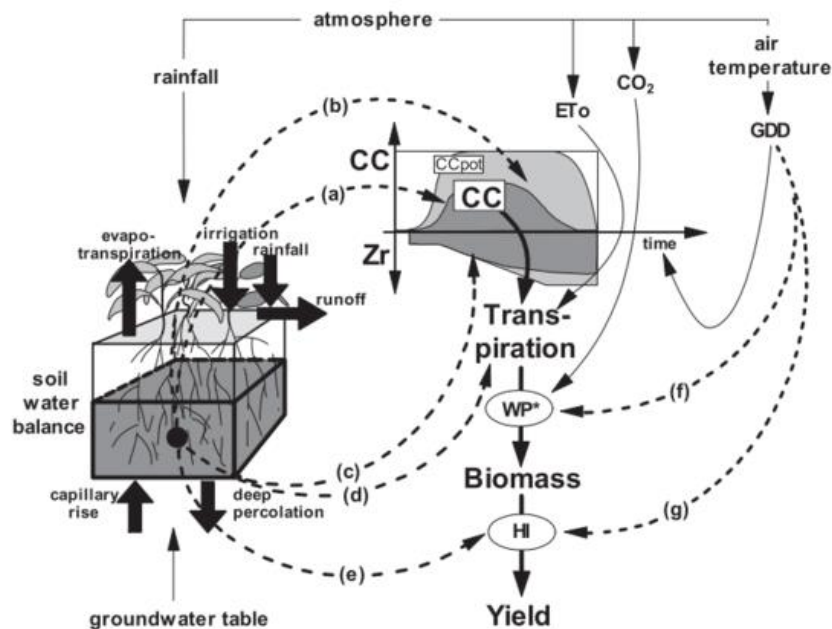


Figure 1: Physiological processes in the AquaCrop model (Vanuytrecht et al., 2014)

In the simulation of crop development, Aquacrop expresses foliage development in terms of FCover, which is alternatively referred to as canopy cover (CC) in AquaCrop. The CC values range

from 0 (time before emergence) to the maximum canopy cover (CC_x), which could be 100% depending on the type of crop under investigation (Vanuytrecht et al., 2014). The development of the canopy is modelled using the initial canopy cover (CC_0) and the CC after emergence (CC_x), which varies depending on the crop and canopy growth coefficient (CGC) (Vanuytrecht et al., 2014). Figure 2 shows the trajectory of the expected canopy cover development stages throughout the crop growth cycle under ideal conditions, i.e., without water stress.

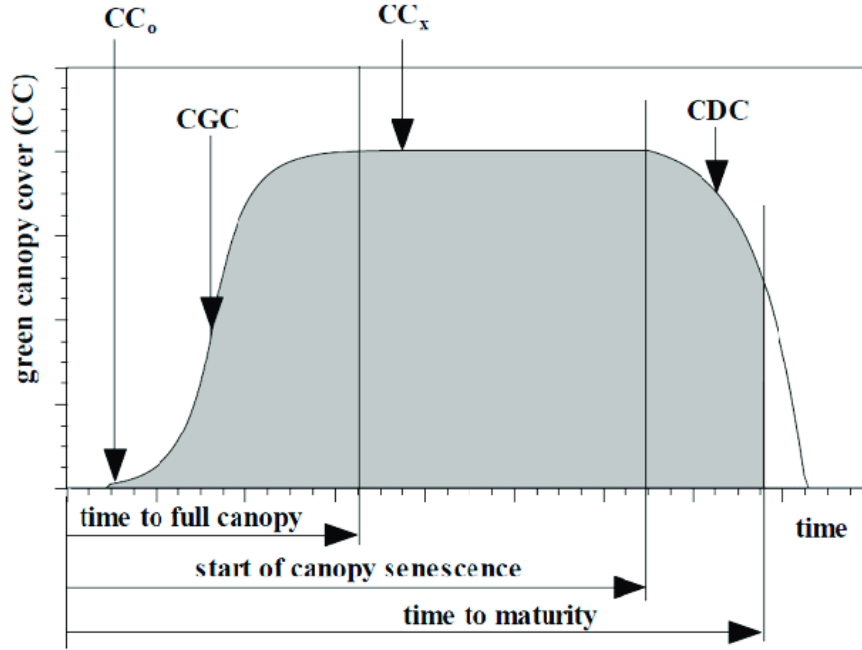


Figure 2: Variation of green canopy cover throughout the growing cycle under non-stress conditions (FAO, 2022)

After reaching the CC_x , the $FCover$ begins to decline because the plant has reached maturity and is ready to be harvested, commonly referred to as senescence. AquaCrop simulates crop development in growing degree days (GDD), meaning that canopy variation is strongly influenced by temperature (Vanuytrecht et al., 2014). Equation 1, as declared by Vanuytrecht et al. (2014), shows the formula for calculating GDD:

$$GDD = T_{avg} - T_{base}$$

Equation 1: Growing degree days equation

Where:

- T_{avg} is the mean of the daily maximum and minimum temperature.
- T_{base} is the temperature below which crop development cannot progress.

AquaCrop also considers Tupper, the temperature above which crop development does not progress. With the estimation of CC and taking the weather data into account, crop transpiration (Tr) is hence calculated by multiplying the potential evapotranspiration (ETo) by the crop coefficient (KcTr) using Equation 2 as declared by Vanuytrecht et al. (2014).

$$Tr = K_s (K_{cTr} \times CC^*) ETo$$

Equation 2: Crop transpiration equation

Where:

- Ks is the stress coefficient.
- CC* is the adjusted FCover.
- (KcTr, x CC*) represents the crop coefficient which is a characteristic of the crop.
- ETo is the potential evapotranspiration.

After establishing crop transpiration, AquaCrop considers additional parameters, i.e., temperature stress coefficient for biomass (Kbs) and normalized water productivity (W.P.*), to simulate the above-ground biomass (AGB) as shown in Equation 3 as utilised by Vanuytrecht et al. (2014) by accumulating the daily biomass production corresponding to ETo.

$$B = K_{bs} WP^* \Sigma (Tr / ETo)$$

Equation 3: Above-ground biomass equation

Establishing the AGB provides an opportunity to calculate the yield. Crop yield (Y) is therefore calculated as a product of AGB and actual harvesting index (HI.). The HI is obtained by adjusting the reference HI., a fraction of the biomass under no-stress conditions, with a stress factor (fHI). Hence, Y is calculated as shown in Equation 4, as declared by Vanuytrecht et al. (2014).

$$Y = fHI \cdot HI_o \cdot B$$

Equation 4: Yield calculation equation

AquaCrop is an appropriate tool for simulating crop productivity in arid and semi-arid areas due to its ability to account for the impact of water scarcity, soil properties, and other environmental factors on crop growth (Wellens et al., 2022). Furthermore, the Aquacrop model has been validated

for various crops in various ASAL areas and has displayed high accuracy and reliability in simulating crop yields under harsh environmental conditions (Ahmadi et al., 2022).

According to a study by Todorovic et al. (2009), field experiments were performed in a typical Mediterranean area of Southern Italy in 2005 and 2007. The experiments assessed AquaCrop, CropSyst, and WOFOST in simulating sunflower growth under different water regimes, i.e., rainfed and irrigated conditions. The models were calibrated using data acquired from fully irrigated fields in 2007. The models compared in terms of yield as follows; AquaCrop yielded an accuracy of $RMSE=0.70$, WOFOST's accuracy was $RMSE=0.79$, and CropSyst's yield accuracy was $RMSE=0.94$ (Todorovic et al., 2009). The lack of the plant physiological information to account for water stress in AquaCrop was one factor that impacted its performance in simulating the yield estimation in the study (Todorovic et al., 2009).

2.4. Assimilation of biophysical variables

Data assimilation is one of the methods known to reduce crop yield model errors. The technique entails using observed or satellite-derived data to update model variables or to estimate model parameters (European Commission. Joint Research Centre. Institute for Environment and Sustainability. and CCAFS., 2012). There are several methods through which model assimilation is undertaken, through updating strategies, calibration or forcing (Dhakar et al., 2022). Figure 3 shows a schematic illustration of the different assimilation strategies of remotely sensed biophysical variables usually utilised in crop yield estimation models.

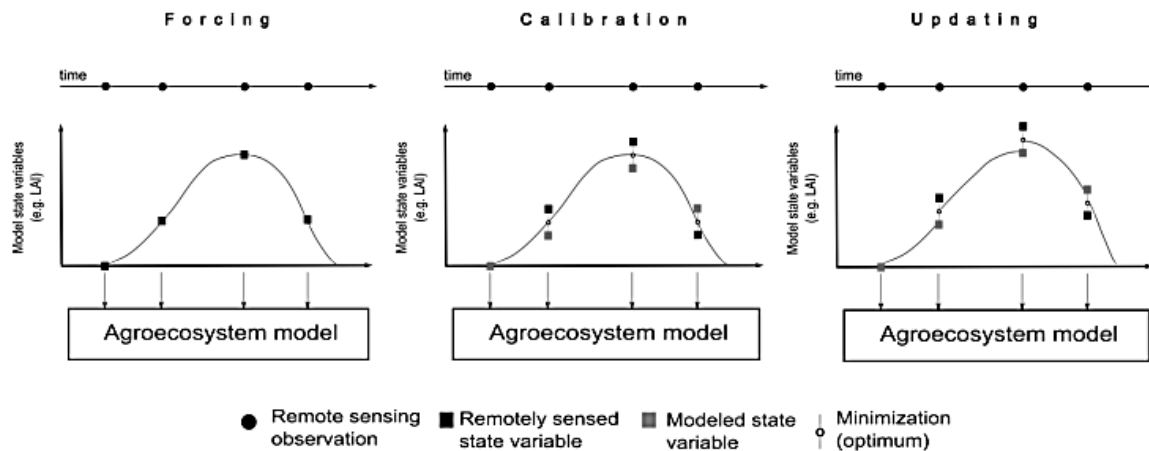


Figure 3: Schematic illustration of the different methods for the assimilation of remotely sensed biophysical variables in crop yield estimation models (Delécolle et al., 1992)

Forcing is a mechanism that directly replaces the model state variable with satellite-derived biophysical variables (Dhakar et al., 2022). Apart from forcing, assimilation algorithms are also used, such as the Kalman Filter (K.F.), Extended Kalman Filter (EKF), Ensemble Kalman Filter (EnKF) and Particle Filter (P.F.). The algorithms derive a weightage between the simulated model state and the observations to minimize the model's uncertainties (Dhakar et al., 2022).

Biophysical variables retrieved from satellite, such as LAI, soil moisture and FCover, can be assimilated into crop models to provide close to accurate information for crop yield estimation at various scales (European Commission. Joint Research Centre. Institute for Environment and Sustainability. and CCAFS., 2012). A study was done by Jin et al. (2020) to assess the impact of the assimilation of FCover in Aquacrop. The study was undertaken through field experiments in 2012 and 2013 at the Xiaotangshan precision agriculture experimental site in Beijing, China using Aquacrop to estimate maize yield. The FCover was derived using several indices obtained from hyperspectral data, and the index with the best fit of FCover was selected for the assimilation process. The estimation of maize yield was significantly improved through the assimilation of FCover using the particle swarm optimization (PSO) algorithm approach in Aquacrop, achieving an accuracy $R^2 = 0.78$ at a regional scale.

A different study by Abi Saab et al. (2021) estimated the yield of winter wheat grown within a Mediterranean environment in Lebanon's Bekaa Valley during three growing seasons ranging from 2017 to 2019 using AquaCrop. Different precipitation patterns characterized the area of interest under the study; hence the model was calibrated for the different water regimes. The FCover was retrieved from Sentinel-2 and assimilated into Aquacrop, and the yield estimation accuracy performance achieved was $RMSE = 0.54$ for rainfed conditions at a regional scale.

In conclusion, previous studies show that accurate and timely crop yield estimates acquired before crop harvesting with the help of improved crop yield estimation models are extremely valuable. In addition, the yield estimation models provide a cheap option for looking into a range of management options apart from yield simulation (Wellens et al., 2013). Accurate yield estimation influences to a great extent the food security, trade and economic policies (Hayes and Decker,

1996) at a regional, national and international level which is of interest to climatologists, farmers, policymakers, planners, practitioners and aid organizations (Araya et al., 2010; Jin et al., 2020).

CHAPTER 3. Research Objectives

The main objective of this study is to assess the improvement of maize crop yield estimation by assimilating remotely sensed biophysical variables in AquaCrop using Western Kenya as a case study. To achieve the main objective, this study seeks to answer the following research question: How does remotely sensed data affect the performance of crop yield estimation models?

The specific objectives are to:

1. Ingest meteorological time series and soil data into AquaCrop and simulate maize yield estimation.
2. Retrieve a biophysical variable, i.e., FCover from Sentinel-2 over the corresponding growing season.
3. Assimilate the earth observation-derived biophysical variables into AquaCrop for the simulated maize yield estimation.
4. Assess and compare the respective yield estimation performances of AquaCrop with and without the assimilation of earth observation-derived biophysical variables.

CHAPTER 4. Materials and Methods

This chapter describes in detail the datasets used in the analysis of this study. In addition, the steps used in pre-processing and processing the dataset are also discussed. It starts by introducing the study area, the methodology workflow, and the tools used in the analysis. Finally, the data sources and composition, including the analysis undertaken, are also discussed.

4.1. Study area

The areas under study are Uasin Gishu and Keiyo Marakwet counties, located in the western part of Kenya within the Rift Valley region. Uasin Gishu county lies between latitude 0°03' N-0°55' N and longitude 34°50' E - 35°37' E, and Keiyo Marakwet County lies between latitude 0°10' N - 1°14' N and longitude 35°12' E-35°42' E (Kenya County Guide, 2023).

Uasin Gishu county has an approximate area of 2955 km² and a population density of 343 people per sq. Km (Kenya National Bureau of Statistics, 2019) and approximately 2100 m above sea level elevation. Uasin Gishu is located within a plateau and has a cool and temperate climate with rainfall ranging between 625 mm to 1560 mm and temperatures between 8.4⁰ C and 27⁰ C (Kenya County Guide, 2023). On the other hand, Keiyo Marakwet, which borders Uasin Gishu to the northeast, has an approximate area of 3,032 Km² with a population density of 150 people per sq. Km (Kenya National Bureau of Statistics, 2019) and an elevation of approximately 2363.34 meters above sea level. Keiyo Marakwet experiences temperatures between 14°C to 24°C and receives rainfall between 400 mm and 1,400 mm annually (Kenya County Guide, 2023).

The two counties are located within the Kenyan agroecological zones, which possess high agricultural potential, and agriculture is the main economic activity of the inhabitants of the two counties (Mcharo and Maghenda, 2021; Pierce, 2016). Maize within the region is grown under small-scale farming and heavily relies on rainfall (Pierce, 2016). Maize is usually planted during the long rainy season, i.e., between March and May, and harvesting occurs from October to mid-December. The study area shown in Figure 4 was adopted due to data availability and the area's suitability to apply the AquaCrop model.

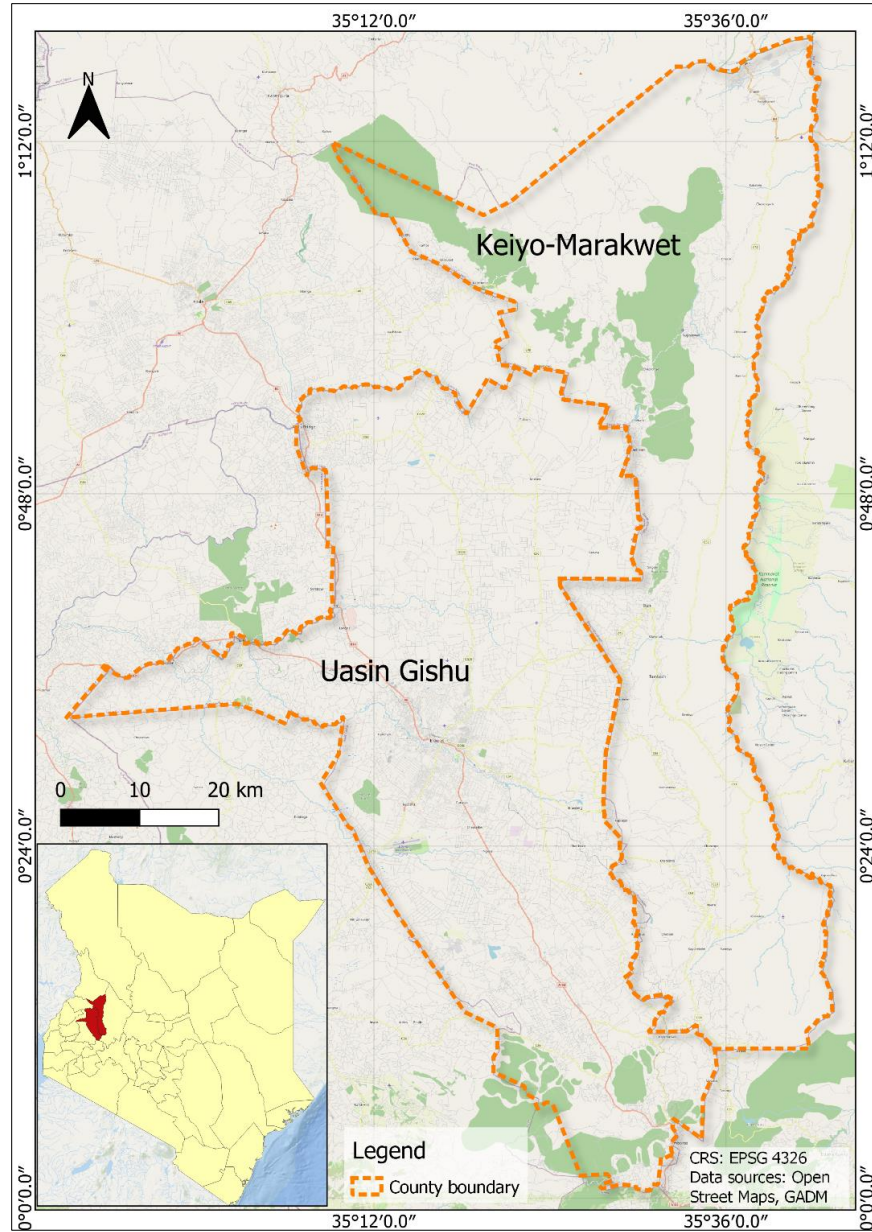


Figure 4: Study area map showing the location of Uasin Gishu and Keiyo Marakwet

4.2. Methodology

Figure 5 shows the general methodology workflow followed to achieve the objectives of this study. The steps highlighted are detailed in the data and processing section. The specific objectives are addressed by the methodology workflow as follows:

- The ingestion of meteorological time series into the AquaCrop model is undertaken by obtaining data from ERA-5 land (ECMWF, 2018) for 2021. The soil dataset is obtained from the Harmonized world soil database (HWSD) (FAO, 2013) and ingested into the model together

with the meteorological time series data. This information helps in yield simulation by the model.

- The FCover data retrieved from Sentinel-2 using the Biophysical variable neural network (BV-net) (Weiss and Baret, 1999) algorithm goes through various pre-processing steps, such as cloud masking and zonal statistics, to obtain the time series profiles for each parcel.
- The resulting FCover time series are used in the process of assimilation in the AquaCrop model, and yield is again simulated.
- The results obtained from the AquaCrop model with and without assimilation are therefore assessed and compared using performance metrics.

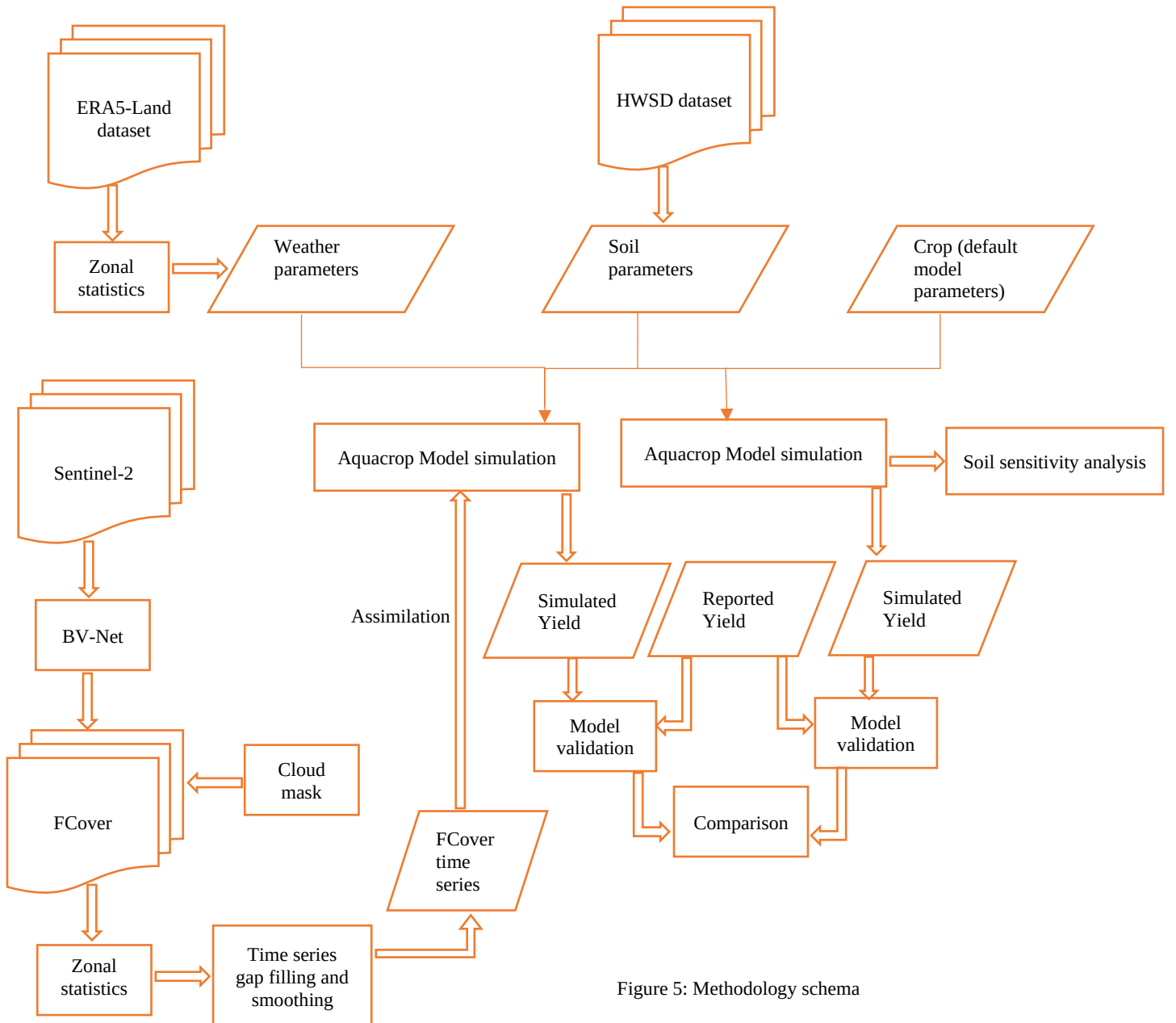


Figure 5: Methodology schema

4.3. Tools

The following platforms were used in processing the dataset: Python Jupyter Notebook v.1(Anaconda, n.d.), AquaCrop-OSPy v.3.7 Google colab Notebook (FAO, n.d.), QGIS version 3.22.12 (Sherman, 2002) and ETo calculator v.3.2(FAO, 2023).

4.4. Data and processing

Table 2 outlines the summary of the data utilised in the analysis. All the datasets collected and used in this study are freely accessible.

Table 2: Dataset summary

Data	Product name	Product type	Resolution Spatial- Temporal	References
Weather data	ERA-5 Land	Maximum temperature, Minimum temperature, Precipitation, Potential evapotranspiration	9km, Hourly, 2021	(ECMWF, 2018)
Soil dataset	HWSD	Soil texture	1km, 1970-2007	(Kenya Soil Survey, 2018), (FAO, 2013)
Satellite data	Sentinel-2	Satellite images	10m, 5 days 2021	(ESA, 2014)
Elevation data	Copernicus Digital Elevation Model	Digital elevation model (DEM)	30m, 2023	(European Space Agency, 2023)
Maize yield data	-	Yield data and parcel boundaries	2021, parcel	(PULA, 2022)

4.4.1. Maize yield dataset

The maize yield field dataset used as the validation data for the model was obtained from Bird's-Eye, a project that collects ground truth georeferenced agriculture-labelled datasets in Kenya (PULA, 2022). The Bird's-Eye project in Kenya is a collaboration between Pula and Tetra Tech and is funded by the Bill & Melinda Gates Foundation (PULA, 2022). The maize yield data was collected by identifying regions similar in rainfall patterns, soil texture, and topography within the Kenyan agroecological zones. The similarity in the characteristics of the chosen areas is important since it ensures the collection of information on crops with similar growth patterns and lengths (PULA, 2022).

The yield in every parcel was determined by placing a single eight by five meters box placement on a selected region within the parcel boundary. Figure 6 shows a box placement exercise within a parcel. The geographical coordinates of the boxes and the parcel boundaries were picked, and this information was saved to aid with quality control and analysis processes afterwards. Analysis and quality assessment are undertaken during the growing phase or two to four weeks before maize harvesting. The quality assessment before harvesting entails ensuring the spatial information related to the parcel and box placements is correct. After harvesting, it is checked that the reported weight of the dry harvest is not more than the fresh harvest (PULA, 2022).



Figure 6: Box placement exercise within a parcel (Bird's-Eye)

Harvesting was undertaken within the box placement, leading to the acquisition of the fresh harvest made up of maize grains and cobs. The fresh harvest is placed in a gunny bag and weighed. Afterwards, the fresh harvest is left to dry, threshed, winnowed, and reweighed, as shown in Figure 7, providing the dry harvest composed of the weight of the grain without the cobs. The weight of the dry harvest was then used to calculate the yield (PULA, 2022).



Figure 7: Drying and weighing of the fresh harvest (Bird's-Eye)

The data acquired from Bird's-Eye used in this study comprised 118 parcels within Uasin Gishu and Keiyo Marakwet whose selection criteria entailed parcels that were homogeneous in terms of purity, i.e., only maize crop was being grown on the parcel, planted maize during the long rain season and produced yield ranging between 2-8.5 t/ha. Figure 8 shows the spatial distribution of the parcels' yield within the two counties showing the relationship between the soil type and elevation to the yield. The maps show that the parcels located within mid-altitude areas with Clay soil produced low yields compared to those in low-altitude regions with Sandy clay loam soil. The elevation dataset used is described in detail in Figure 43 under Appendix 2.

The long rain season begins in March, and some farmers take advantage and undertake their planting during the start of the rains. In March, the rain is not very intensive and intensifies within mid-April; hence, other farmers wait for that period to undertake their planting. Due to this

difference in planting dates decision made by the farmer, the dataset contains varying planting and harvesting dates. Table 3 shows the distribution of planting dates for the selected parcels within the two planting months, i.e., March and April.

Table 3: Parcels' distribution within the planting and harvesting months

Planting months	Harvesting months	Number of Parcels
March	October – December	40
April	October – December	78

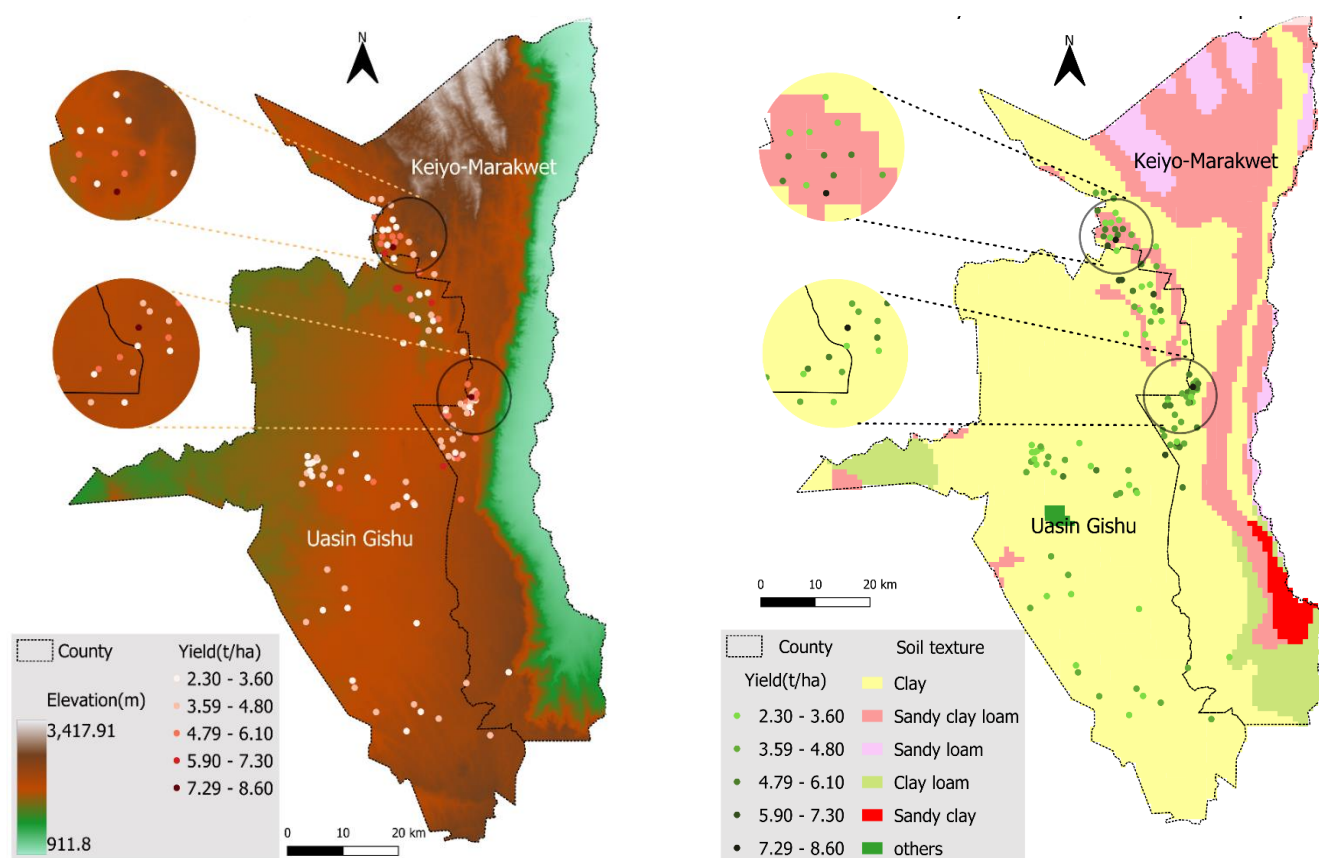


Figure 8: Yield distribution with respect to elevation and soil texture

4.4.2. Parcel data verification

The quality control of the parcels' dataset was undertaken through two steps. First, the yield values were evaluated, and a threshold of yield values above 2t/ha was set as acceptable for further analysis. Secondly, due to the range of distribution of the reported yield within the parcels in the two counties, it was necessary to investigate the heterogeneity of the parcels to establish the reason behind the variation in the reported yield. In addition, it was also essential to check if the reported yield corresponds to the satellite's maximum FCover values. The standard deviation was the measure used to check the heterogeneity of the parcels. Standard deviation is a measure of how dispersed the values are from the mean and is computed as shown in Equation 5.

$$Sd = (\sum (x_i - \bar{x})^2 / (n-1))^{1/2} \quad \text{Equation 5: Standard deviation equation}$$

where:

- x_i is the value of a point in the data.
- $\bar{x}$ is the mean value of the data.
- n is the number of values in the data.

4.4.3. Satellite dataset

Sentinel-2 datasets LIC were acquired from the Copernicus Open Access Hub for January 2021 to December 2021, covered by tiles 36NYF and 36NYG. The acquisition had a 5-day interval corresponding to the revisit time of Sentinel-2 within the equatorial regions. The Sentinel-2 dataset was important as it is used to obtain the FCover time series utilised in this study's assimilation process.

4.4.4. FCover processing

The U.C Louvain Geomatics lab processed the FCover used in the analysis. First, the L1C Sentinel-2 time series dataset for 2021 covering the area of interest was obtained. The L1C dataset was pre-processed using Sen2cor to obtain the L2A dataset. L1C data products provide the top of the atmosphere reflectance. By pre-processing the L1C data to remove the effect of the atmosphere, L2A data was obtained, which provides the bottom of the atmosphere reflectance. The Resulting L2A images were then used to compute the FCover through the BV-Net algorithm of the Sen2Agri system. The function of BV-Net is based on undertaking the simulation of biophysical variables

by applying the inversion of PROSPECT and SAIL(PROSAIL) radiative transfer models (Dhakar et al., 2022). BV-Net is an Artificial neural network (ANN) algorithm (Weiss and Baret, 1999) which has been calibrated in Belgium on wheat fields providing good results for other locations (Delloye et al., 2018).

4.4.5. Cloud masking

The resulting FCover produced by ANN contained cloud cover. The cloud cover must be masked to prevent wrong FCover values from being retrieved within the given parcels. Cloud masking was undertaken using FMask with a threshold of 10%, which the U.C Louvain Geomatics lab provided alongside the FCover dataset. The cloud mask detected clouds, cloud shadows, and water bodies such as rivers and water bodies with vegetation and land surfaces. Raster algebra within Jupyter Notebook using Python was used to subtract the pixels with cloud and cloud shadow from the FCover raster using the FMask raster. Figure 9 shows the features detected by FMask for one of the images used.

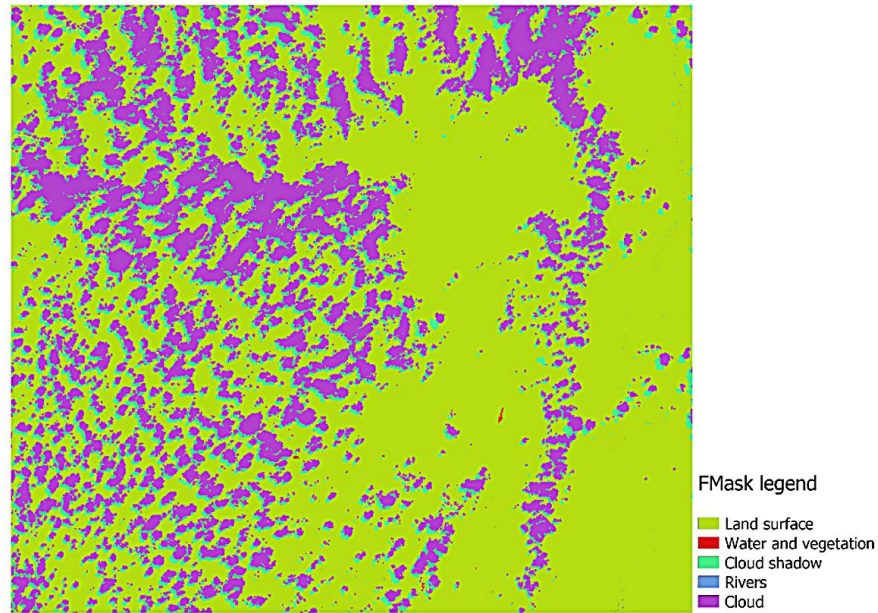


Figure 9: Features detected by FMask

4.4.6. FCover values extraction, gap filling and smoothing

After the cloud masking process, the mean FCover values were extracted through zonal statistics in Python for all the parcels. Before employing zonal statistics, an internal buffer of 5m was applied to the parcel boundaries to avoid boundary edge noise related to mixed pixels, including features such as roads from interfering with the retrieved FCover values for a given parcel. The presence

of cloud cover caused a hindrance in acquiring data for all the satellite revisit times. Hence, it resulted in missing FCover values for some dates after applying the cloud mask to the FCover images. Therefore, to obtain a continuous profile for the FCover throughout the year, gap-filling and smoothing of the data points were undertaken using the Stavisky-Golay approach.

Savitzky-Golay is an algorithm used in signal processing to eliminate outliers in a temporal signal and improve the smoothness of the signal based on least squares approximation (Niedźwiecki et al., 2021). The advantage of applying Savitzky-Golay is that it increases the precision of the data without distorting the signal. The algorithm calculates a polynomial fit using polynomial degree and window size. This study's analysis used the polynomial degree of 2 and the window size of 20.

4.4.7. Weather dataset

The daily meteorological dataset for 2021 that was used as input in the model was obtained from ERA-5 land. ERA-5 land provides hourly weather data from 1950 at a spatial resolution of 9km grid spacing (ECMWF, 2018). The data obtained included minimum and maximum temperature, rainfall, and potential evapotranspiration. Figures 10, 11 and 12 show an example of the temperature, precipitation and potential evapotranspiration data profiles obtained for one of the parcels under investigation.

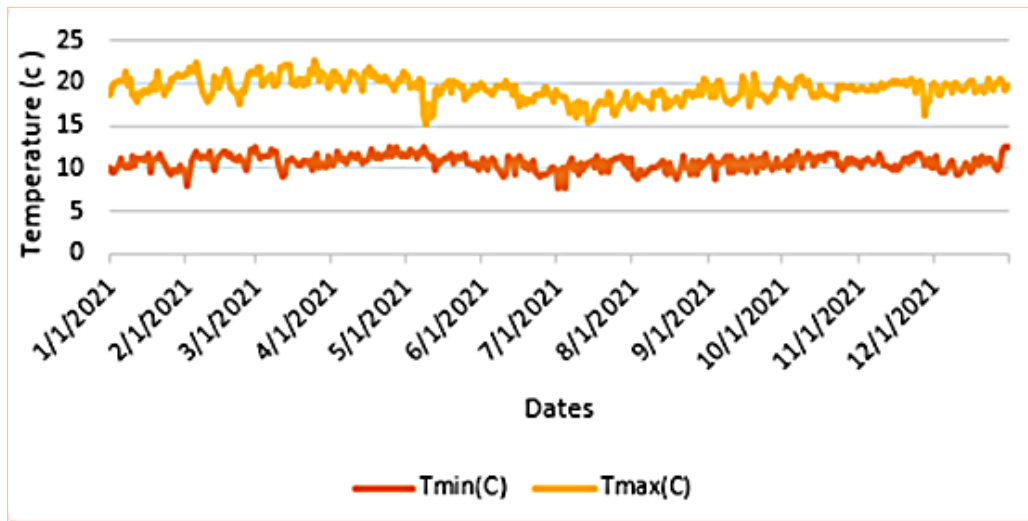


Figure 10: Temperature data profile for parcel code_0 in 2021

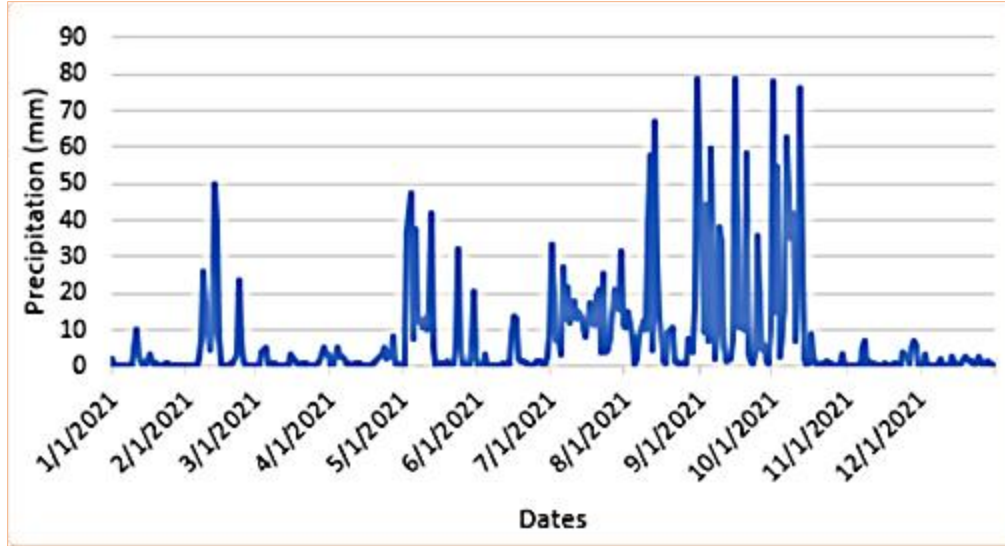


Figure 11: Precipitation data profile for parcel code_0 in 2021

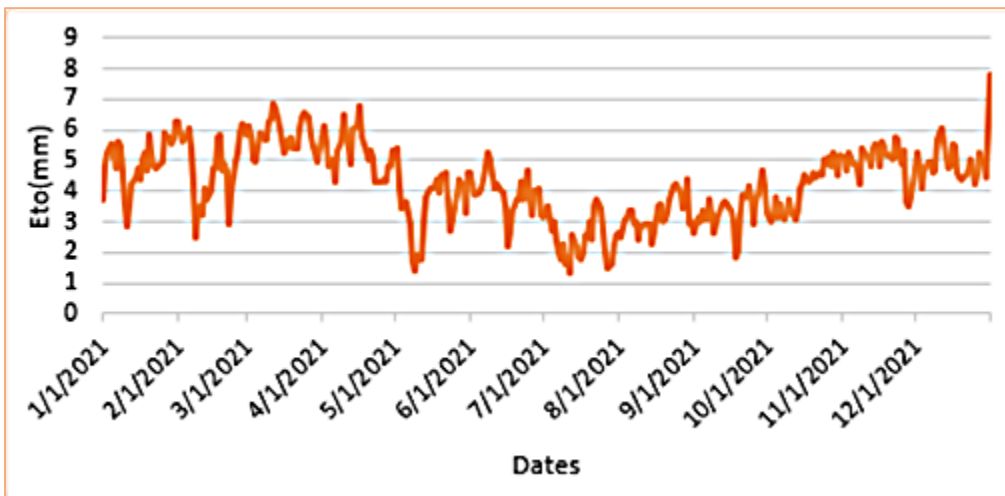


Figure 12: Potential evapotranspiration data profile for parcel code_0 in 2021

4.4.8. Weather data processing

The weather dataset obtained from ERA-5 land was in NetCDF format, which was processed in Python using zonal statistics based on the coordinates of the parcels to derive the mean daily weather information throughout the year. Additional processing steps were required to convert the data retrieved from ERA-5 land into a Txt file format acceptable by AquaCrop.

Temperature information from the ERA-5 land was provided in Kelvins, which was converted to degrees Celsius. Precipitation and potential evapotranspiration were converted from meters to

millimetres. Potential evapotranspiration retrieved from ERA-5 land had negative values. Verification was undertaken to determine whether the Eto dataset was valid by deriving the potential evapotranspiration for the parcels using the FAO ETo calculator. ETo calculator utilises the Penman-Monteith equation (FAO, 2009) given by Equation 6.

$$ETo = (0.408 \Delta (Rn - G) + \gamma (900/T + 273) u_2 (e_s - e_a)) / \Delta + \gamma (1 + 34.0 u_2) \quad \text{Equation 6: Penman-Monteith equation}$$

Where:

- ETo reference evapotranspiration [mm day⁻¹],
- Rn net radiation at the crop surface [M.J. m⁻² day⁻¹],
- G soil heat flux density [MJ m⁻² day⁻¹],
- T mean daily air temperature at 2 m height [°C],
- u₂ wind speed at 2 m height [m s⁻¹],
- e_s saturation vapour pressure [kPa],
- e_a actual vapour pressure [kPa],
- e_s-e_a saturation vapour pressure deficit [kPa],
- Δ slope vapour pressure curve [kPa °C⁻¹],
- γ psychrometric constant [kPa °C⁻¹]

ETo obtained through the ETo calculator summed up to approximately a positive value of the potential evapotranspiration obtained from ERA-5 land but slightly lower, probably due to the lack of precise dataset to use as input in the ETo calculator such as radiation, wind speed and humidity. Hence, the potential evapotranspiration obtained from ERA-5 land was multiplied by -1 and utilised as input in the model.

The retrieved weather information was then stored in different files containing each parcel's maximum temperature, minimum temperature, potential evapotranspiration, and precipitation. The resulting files were formatted following the recommended column names in AquaCrop, i.e., Day, Month, Year, Tmin(C), Tmax(C), Prcp(mm), Eto(mm) and converted into a text file.

4.4.9. Soil texture dataset

The soil texture dataset was obtained from HWSD, which provides the soil dataset in raster format alongside an access database containing soil information. Information obtained from HWSD is based on the Soil and Terrain database for Kenya (KENSOTER) dataset, which was last updated in 2007. The KENSOTER dataset was compiled in 1996 by the Kenya Soil Survey, based on the Exploratory Soil Map and the Agro-Climatic Zone Map of Kenya (Kenya Soil Survey, 2018). KENSOTER provides information on soil, i.e., bulk density, soil profiles, pH, salinity, texture, water holding capacity, nutrients etc. The soil type within the area of interest was classified using the soil texture triangle according to U.S. Department of Agriculture (2017), shown in Figure 13. The soil dataset was important as it was used to define the soil texture within each parcel.

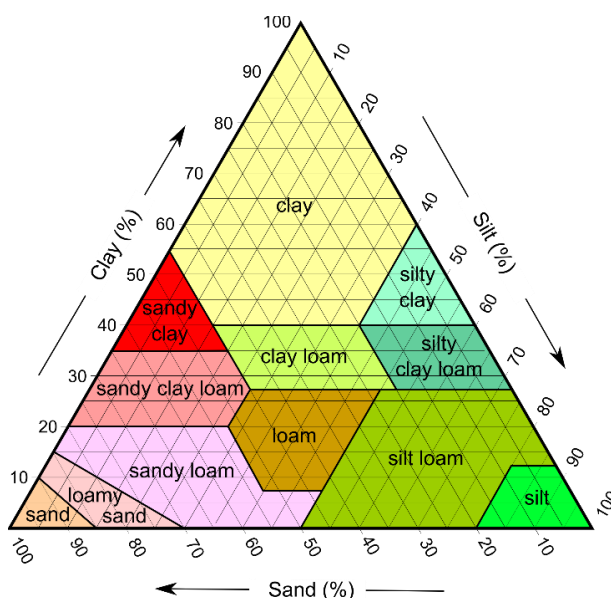


Figure 13: Soil texture triangle defined by USDA

4.4.10. Soil data processing

The soil dataset was processed within the QGIS environment. First, the raster dataset was clipped to the area of interest, and the resulting layer was vectorized. Second, the vector layer's attribute table was joined to the imported dataset obtained from the access database provided alongside the dataset. The access database contained information on soil type, and this field was used in classifying the raster layer according to the soil texture information after the join was undertaken. Using the soil texture triangle defined by USDA, the soil type layer was classified according to international standards (see Figure 45 under Appendix 4), and the resulting map is shown in Figure 14. Finally, an intersection between the soil layer and the parcel was undertaken to obtain the soil

texture within each parcel. The soil map shows that the study area is predominantly covered by Clay, Clay loam and Sandy clay loam soil textures, with clay soil covering a large proportion of the area.

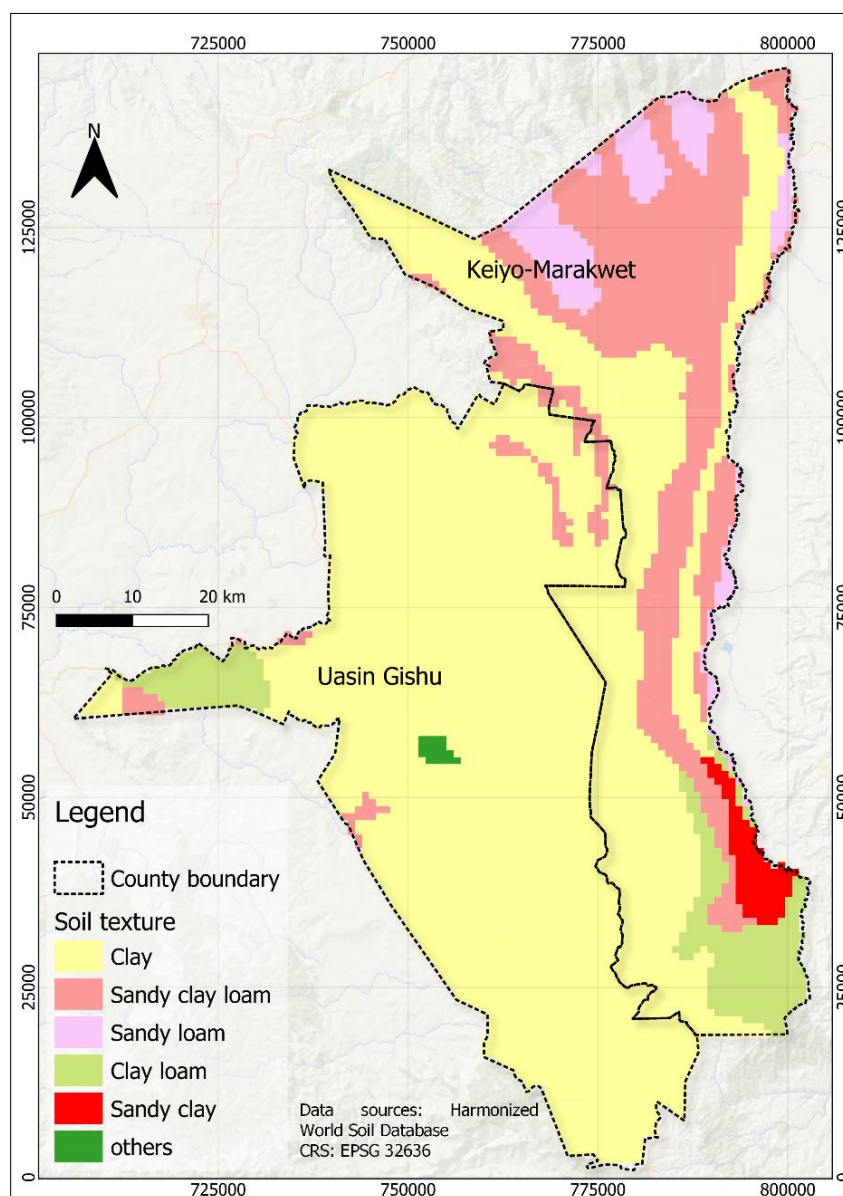


Figure 14: Map showing the distribution of soil textures in Uasin Gishu and Keiyo Marakwet

4.4.11. Aquacrop data and yield simulation

The Aquacrop model has specific parameters that must be defined before using the model to simulate yield. The list of parameters specified from reference papers used in this study is indicated

in Table 4, which includes the base and the upper cut-off temperature. The default model values were used for parameters such as maximum FCover, emergence, CDC and CGC in performing the simulation and obtaining maize yield values for all the parcels.

Due to the lack of data on the initial water content for the parcels, the values for each parcel were obtained by running a simulation in AquaCrop from January till a day before the planting date to estimate the soil water content. The forcing technique replaced the maximum FCover and the emergence date during the assimilation process. Additional variables, such as crop information, soil type, planting and harvesting dates, and weather, were explicitly defined for each parcel.

Table 4: AquaCrop parameters

Parameter	Value	Reference
Base cut-off temperature	8 °C	(Mohamed Sallah et al., 2019)
Upper cut-off temperature	30 °C	(Mohamed Sallah et al., 2019)
Maximum FCover	45 - 88%	Assimilated
Duration to emergence	13 - 42 days	Assimilated

4.4.12. Model sensitivity analysis

It is essential to test which parameters affect the performance of a model through sensitivity analysis to ensure that those parameters are well calibrated for better performance. The soil parameter's sensitivity was undertaken in this study. Hence, the two soil textures, Sandy clay loam and Clay were used to test the model's sensitivity. The sensitivity analysis was undertaken by setting Clay or Sandy clay loam as the soil type for all the parcels when simulating the yield.

4.4.13. Assimilation of FCover in AquaCrop

FCover expansion implies crop development, and it is therefore taken that maximum FCover indirectly provides useful information about the flowering stage of a crop. Considering the maximum FCover for each parcel's FCover profile can significantly improve crop modelling results (Mohamed Sallah et al., 2019). Additionally, according to Mohamed Sallah et al. (2019), the duration to emergence is also a significant factor to consider, and when it is not well defined,

it may result in a difference between the reported and the simulated crop cycle and also in the final simulated biomass and yield.

The maximum FCover obtained was considered for each parcel, and the emergence date was derived from the FCover profile by calculating the emergence FCover values for all parcels using the same procedure applied by Mohamed Sallah et al. (2019), as shown in Equation 7. The procedure ensures a constant time between emergence and maximum FCover since the CGC is a conservative parameter that cannot be changed for each field but remains unique for each crop cultivar.

$$EFCover = FCover_{min} + Tsos \times (FCover_{max} - FCover_{min}) \quad \text{Equation 7: Emergence FCover value calculation}$$

Where:

- EFCover is the FCover value at emergence.
- FCover_{min} is the minimum FCover value at the start of the season, as defined on the FCover profile.
- FCover_{max} is the maximum FCover value obtained from the FCover profile.
- Tsos is set to a constant of 0.15, as defined by Mohamed Sallah et al. (2019)

The most vital parameters required for assimilation in AquaCrop, according to Mohamed Sallah et al. (2019), are the value of FCover at the start of the season, which is linked to crop emergence, and the maximum FCover.

4.4.14. Model performance evaluation

Model performance evaluation without assimilation and with assimilation of FCover information was undertaken using four metrics, i.e., R^2 , RMSE, mean squared error (MSE), Nash-Sutcliffe Efficiency (NSE) and mean actual error (MAE). The MSE is defined as the average of the squared differences between predicted and actual values in a dataset, and the formula used to calculate it is shown in Equation 8. The values of 0 MSE describes that predicted values are equal to the actual values (Gupta and Kling, 2011).

$$MSE = 1/N \sum (Y - Y_i)^2 \quad \text{Equation 8: MSE equation}$$

Where:

- N is the number of data points.
- Y is the actual value.
- Y_i is the predicted value.

MAE is the average of all absolute errors in a model prediction and is calculated as shown in Equation 9.

$$MAE = 1/N \sum |x_i - x|$$

Equation 9: MAE equation

Where:

- N represents the number of errors.
- $|x_i - x|$ represents the absolute errors.

RMSE is calculated as the square root of the MSE, as shown in Equation 10. RMSE measures the difference between predicted and observed values by taking the square root of the average squared differences. Values of RMSE closer to zero indicate that the predicted values are more comparable to the actual values.

$$RMSE = (MSE)^{1/2}$$

Equation 10: RMSE equation

The R^2 value is the coefficient that shows the goodness-of-fit of a model Equation 11 shows the formula used in the calculation. R^2 values close to 1 indicate that the simulated values are close to the actual value.

$$R^2 = \sum (Y_j - Y_i)^2 / \sum (Y - Y_i)^2$$

Equation 11: R^2 equation

Where:

- Y_j is the predicted value.
- Y_i is the average of the observed value.
- Y is the observed value.

NSE is a measure that compares the predicted value with the observed value along the 1:1 line to determine how well a given simulation model performs. NSE is based on the Nash-Sutcliffe equation, which considers the biases and predicted values variances (Lyle et al., 2013). The larger the NSE value, the better the model prediction performance. NSE values range from $-\infty$ to 1, with a negative value implying a significant systematic error in the model simulation (Gupta and Kling, 2011). NSE was calculated using Equation 12.

$$\text{NSE} = 1 - (\sum(X - X_j)^2 / \sum(X - X_i)^2) \quad \text{Equation 12: NSE equation}$$

Where:

- X_j is the predicted value.
- X_i is the average of the observed value.
- X is the observed value.

To compare the model performance without and with the assimilation of FCover, a null hypothesis was undertaken to check if the difference in the performance is statistically significant. Equations 13 and 14 were used as developed by Fisher and R.A. (1921). Fisher exact test is a statistical test employed to assess whether there is a significant difference between the proportions of categories.

$$r' = (0.5) \log_e |(1+r)/(1-r)| \quad \text{Equation 13: Correlation coefficient equation}$$

$$z = (r'_1 - r'_2) / ((1/n_1 - 3) + (1/n_2 - 3))^{0.5} \quad \text{Equation 14: Test statistics equation}$$

Afterwards, the p-value was computed using the z obtained. If the p-value was less than 0.05 using the two-tail test with 95% confidence, r_1 would be considered statistically significant than r_2 .

4.4.15. Data aggregation

The yield information obtained was at a parcel level; hence, it was essential to aggregate the data to a regional level to assess the model performance at a different scale. The data size of 108 parcels only made it possible to aggregate at the ward level with 15 parcels per ward. In cases where the number of parcels per ward was few, wards located next to each other possessing similar elevation and soil characteristics were merged. Equation 15 shows the formula used in aggregation as referenced from the study of Porwollik et al. (2017). The map in Figure 15 shows the distribution of the parcels considered for aggregation within the wards of Uasin Gishu and Keiyo Marakwet.

$$\text{Yield_agg(t/ha)} = (\sum V_i * A_i) / \sum A_i$$

Equation 15: Yield aggregation calculation equation

Where:

- V_i is the simulated or reported yield per parcel.
- A_i is the area of the parcel.

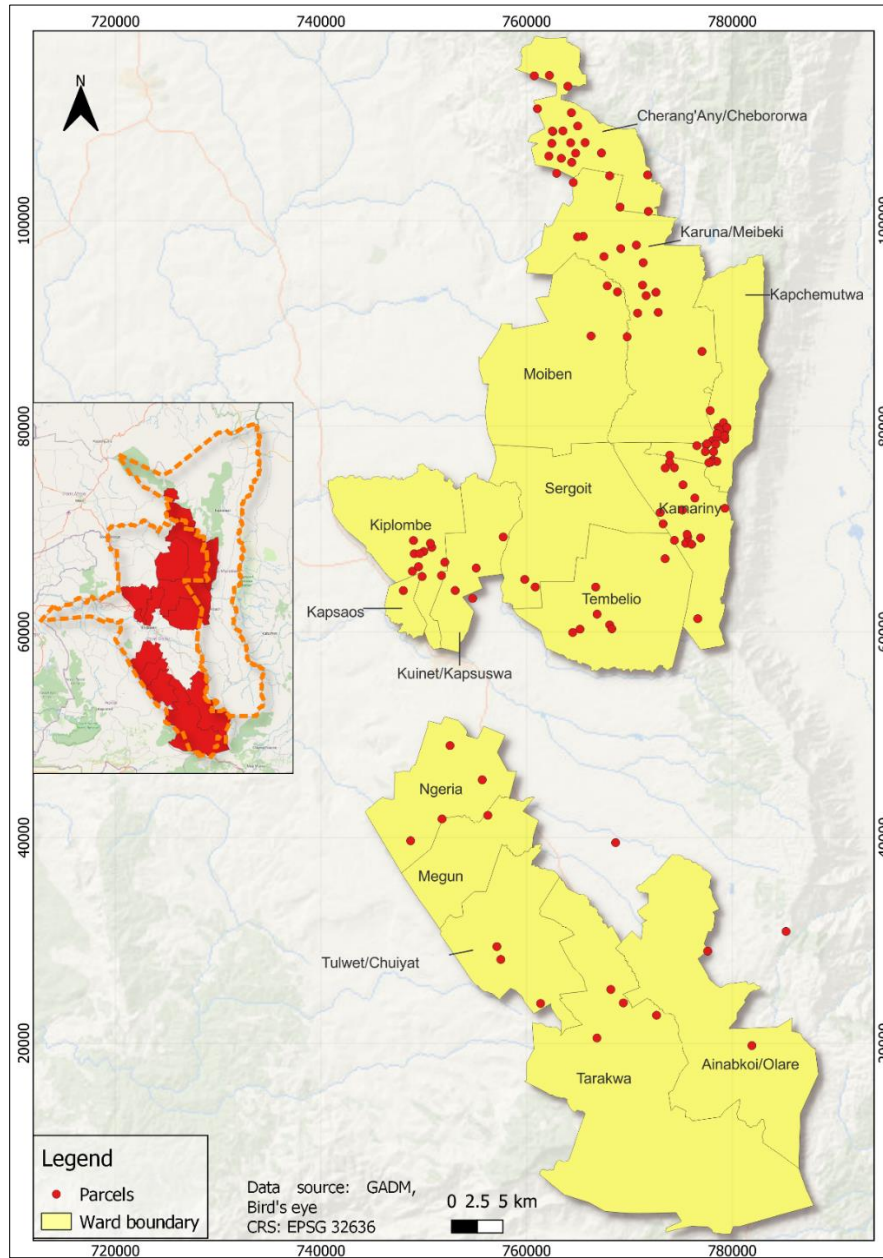


Figure 15: Uasin Gishu and Keiyo Marakwet Wards considered for yield aggregation

CHAPTER 5. Results and Discussion

This chapter delves into the results obtained from the previous section. The results from verification and analysis are presented, including AquaCrop's performance at different scales. The results are then further discussed, deriving knowledge from the information obtained through analysis. Finally, an in-depth analysis comparing the outcome of this study to others is also undertaken.

5.1. Results

5.1.1. Parcel yield verification results

Figure 16 shows the histogram of the yield distribution produced in 2021 by the parcels selected for this study. For the selected dataset, 102 parcels within the two counties produced a yield ranging between 2-5.5 t/ha, while 16 parcels produced a yield ranging between 5.6-8.5 t/ha.

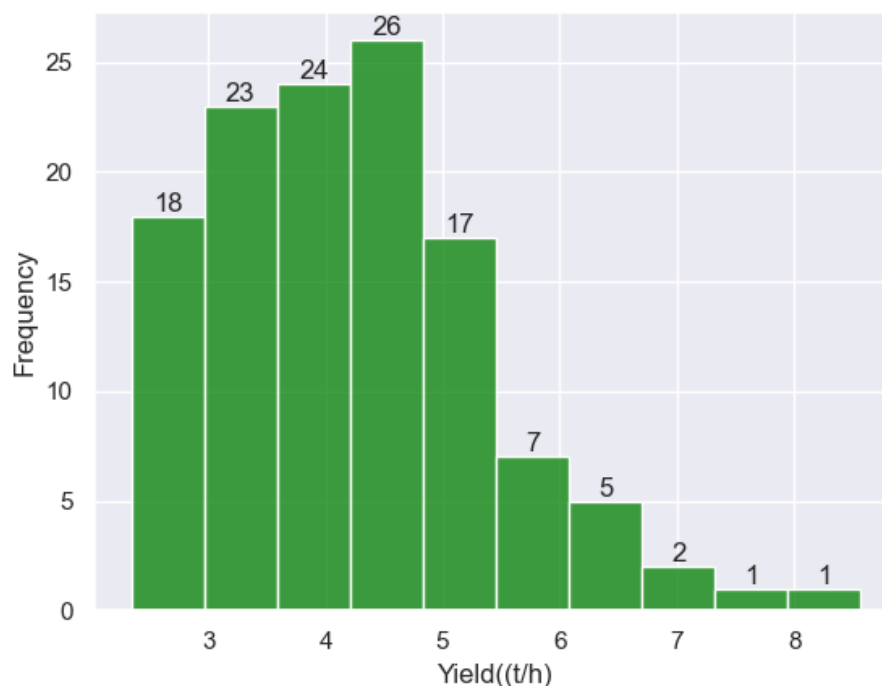


Figure 16: Distribution of maize yield for selected parcels in 2021

Considering the standard deviation of FCover, each parcel's heterogeneity level was determined. Figure 17 shows a scatter plot indicating the variation of heterogeneity in the parcels under investigation. Taking the heterogeneity threshold as 0.1 due to the size of the dataset, 10 parcels were considered highly heterogeneous. Parcel code_73, with a heterogeneity value of 0.19 and a

reported yield of 4.15 t/ha, is shown in Figure 18 overlaid over the FCover and a True colour composite image. The True colour composite shows the FCover values around the top left corner of the parcel being lower than the rest of the parcel. The difference in FCover values indicates that if the box placement was undertaken around the parcel's top left corner, the reported yield would have been lower than if it was placed in another section of the plot.

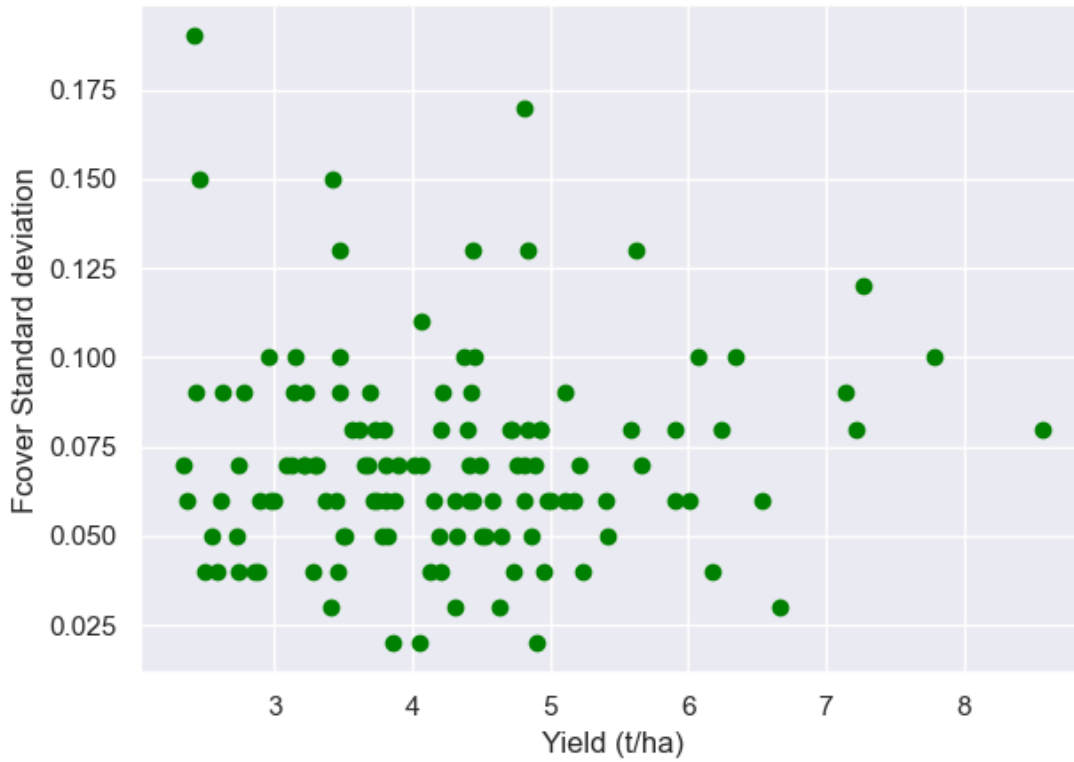


Figure 17: Parcels' heterogeneity

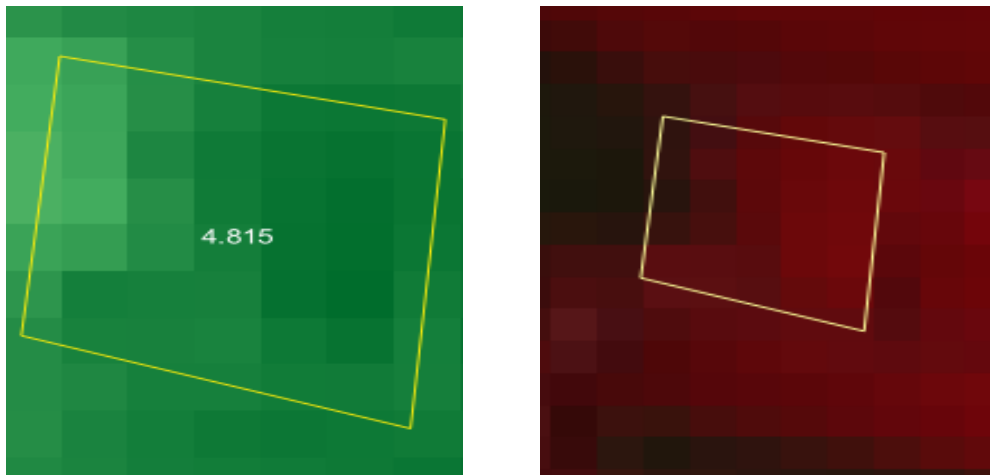


Figure 18: FCover and True colour composite of parcel code_73 showing different vegetation density

5.1.2. Parcel yield analysis results

Analysis of the parcel dataset investigated the relationship between crop cycle duration and planting dates. Figure 19 shows parcels with long crop cycle duration had a planting date in early March, and short crop cycle durations were recorded for parcels with a planting date between late March and April. Additionally, the relationship between the maximum FCover for every parcel and the soil texture showed that some parcels with Sandy clay loam soil produced higher maximum FCover values than the Clay soil, as shown in Figure 20. However, most parcels' maximum FCover values lie within the same range regardless of the soil texture. Furthermore, elevation also played a role in the crop cycle duration. Figure 21 shows the relationship with parcels located within high-elevation areas having longer crop cycle duration.



Figure 19: Crop cycle duration against planting dates



Figure 20: Maximum FCover by soil texture

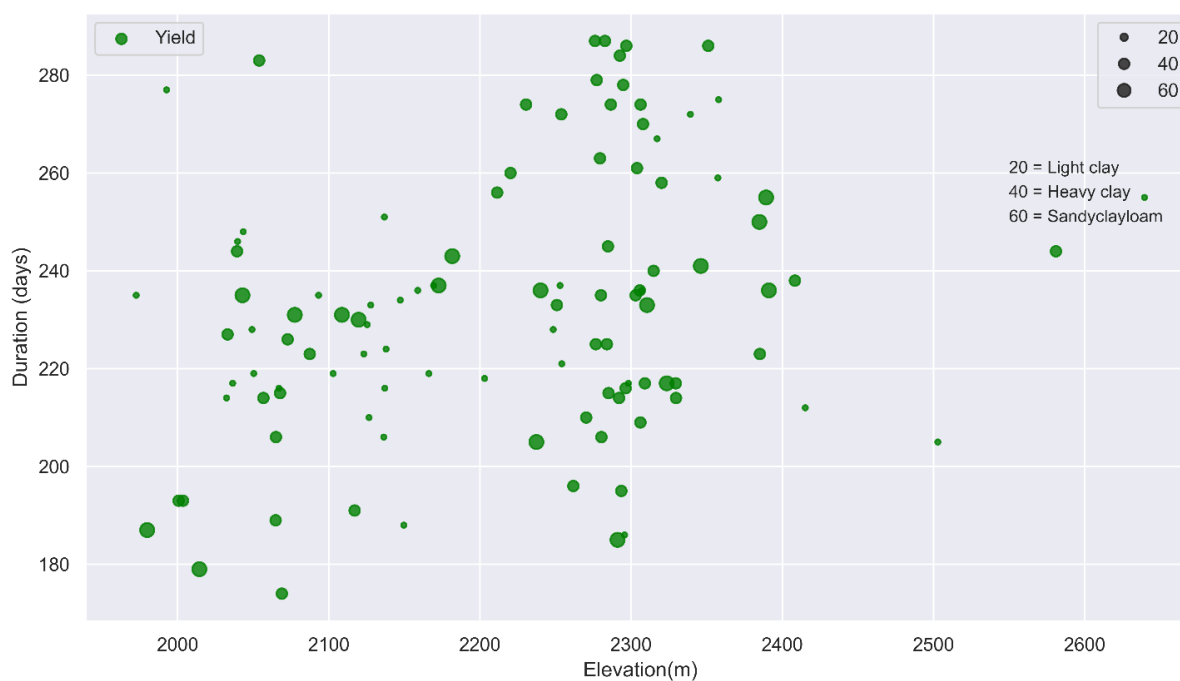


Figure 21: Relationship between elevation and crop cycle duration

5.1.3. FCover values

Figure 22 shows examples of results obtained through the process of Savitzky-Golay. The profiles also show the values of maximum FCover indicated by the orange line obtained and intended to

be used in the model's assimilation process. FCover profiles for the additional 116 parcels are shown in Appendix 1. Some of the parcels' FCover time series did not follow an ideal phenological profile due to the presence of clouds that resulted in missing a significant number of valid data points, only resulting in outliers, as shown in Figure 23; hence, these parcels were excluded from the analysis process.

The FCover profiles, as shown in Figure 22, were used in obtaining the crop emergence date. Figure 24 shows the point taken as emergence dates for some parcels derived from the FCover profiles, corresponding to 34 and 30 days, respectively. The information derived from the FCover profile, the maximum FCover and the time between the start of the season and emergence were assimilated through forcing by directly replacing the default model parameters. Table 5 shows sample data of the information used in the assimilation process for the 20 parcels taken as the test sample before applying the same procedure to the rest. The parcels were selected, taking into account the representation of the different variations in factors within the parcel dataset. Factors such as the soil texture, maximum FCover, reported yield and planting dates were used to select the test parcels but excluded parcels with high heterogeneity.

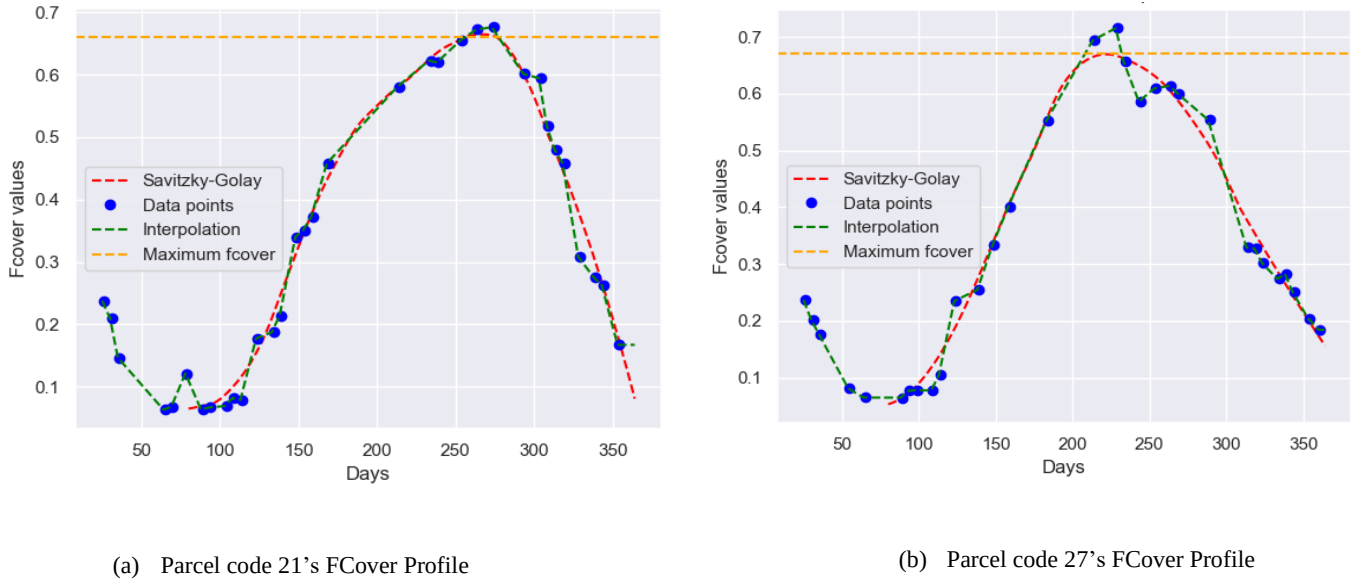


Figure 22: Parcels' FCover profiles with the orange line representing the maximum FCover values

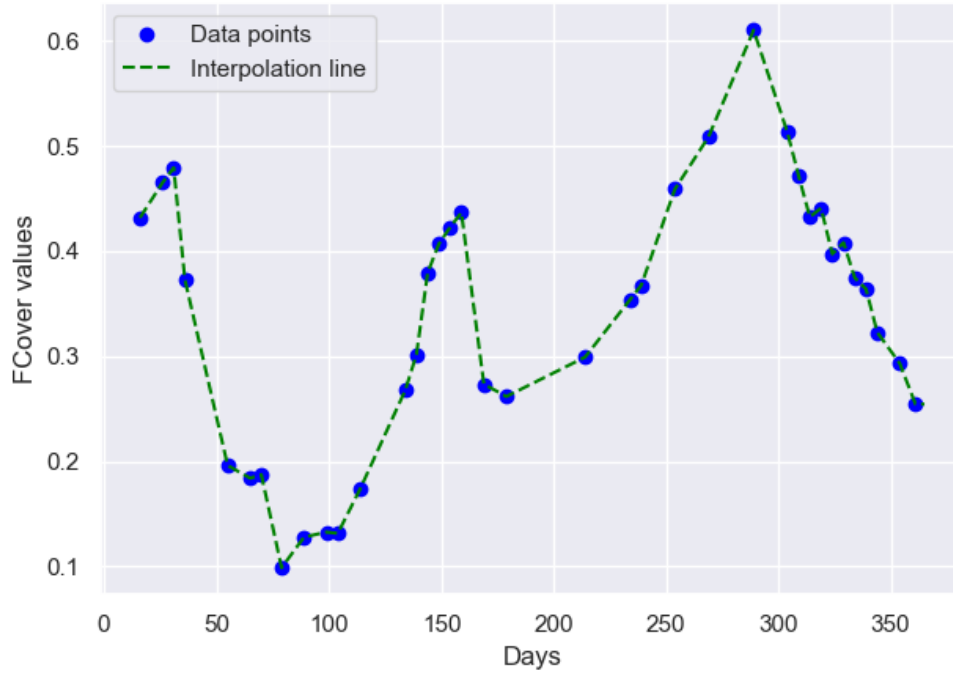
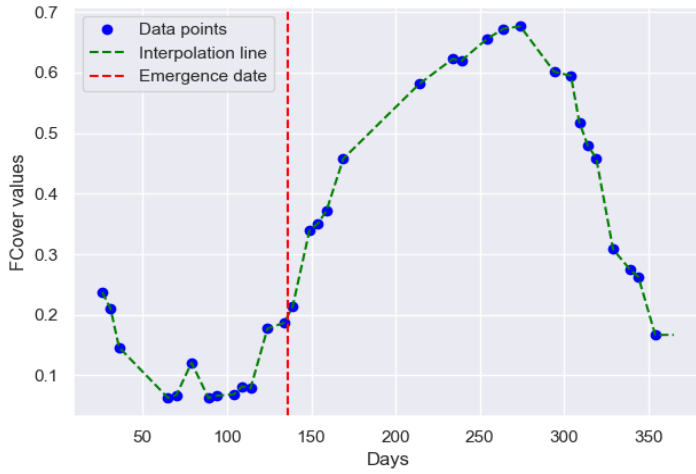
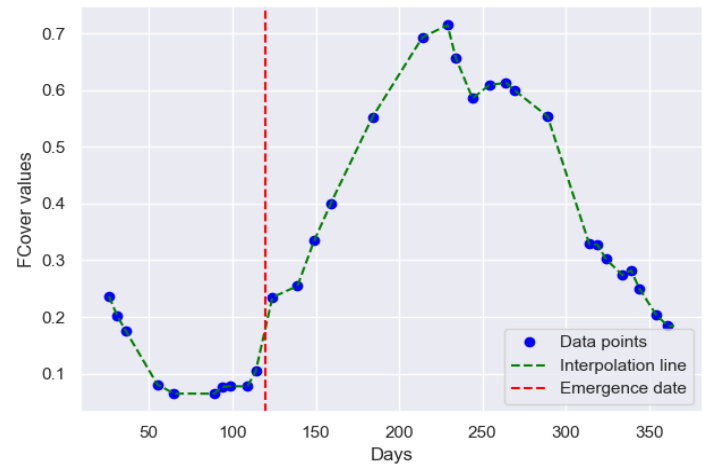


Figure 23: Parcel with an undesirable FCover time series



(a) Parcel code 21's FCover profile



(b) Parcel code 27's FCover Profile

Figure 24: FCover profiles with the red line showing the emergence date

Table 5: Test dataset containing Maximum FCover and duration between crop emergence and maximum FCover

Parcel Code	Planting date	Yield	Maximum FCover values	Emergence date	Soil type	Crop cycle duration
143	4/18/2021	2.6	0.44	5/15/2021	Sandy clay loam	179
77	4/4/2021	7.8	0.73	4/23/2021	Sandy clay loam	230
5	3/8/2021	3.47	0.54	5/17/2021	Clay	287
99	4/5/2021	3.795	0.54	5/23/2021	Clay	219
103	4/3/2021	3.5	0.58	5/21/2021	Clay	217
96	4/18/2021	3.1	0.58	5/29/2021	Sandy clay loam	220
58	4/3/2021	4.9	0.71	5/3/2021	Sandy clay loam	236
62	4/3/2021	4.8	0.71	4/11/2021	Sandy clay loam	235
75	4/2/2021	5.9	0.79	4/9/2021	Sandy clay loam	231
128	3/24/2021	3.8	0.53	5/21/2021	Sandy clay loam	205
2	4/5/2021	3.46	0.55	5/3/2021	Clay	226
18	3/7/2021	4.06	0.67	5/26/2021	Clay	274
27	3/8/2021	4.06	0.7	5/1/2021	Clay	265
32	3/5/2021	4.9	0.68	5/15/2021	Clay	258
51	3/27/2021	3.8	0.58	4/18/2021	Clay	240
77	4/5/2021	7.78	0.73	4/23/2021	Clay	230
129	3/28/2021	3.71	0.57	5/25/2021	Clay	205
60	3/27/2021	3.69	0.67	5/9/2021	Clay	236
134	4/13/2021	6.67	0.69	4/29/2021	Clay	195
104	4/7/2021	7.14	0.64	4/29/2021	Clay	217

5.1.4. Relationship between yield and FCover

The relationship between the maximum FCover retrieved from every parcel's profile and the yield was assessed, as shown in Figure 25. The outcome yielded a R^2 of 0.2732, indicating that there is a moderate correlation ($R= 0.5$) between the Maximum FCover and the yield for the 108 parcels. Due to the difference in the planting dates, the dataset was further divided into three parts, i.e.,

parcels with planting dates falling between the 3rd to 26th of March, the 26th of March to 14th of April and the 15th to 27th of April.

Figure 26 shows a graph of the relationship between the reported parcel yield and planting dates that fall between the 3rd and 26th of March. There seems to be a moderate correlation ($R=0.522$) between the yield and maximum FCover for parcels planted between the 3rd and 26th compared to the graph in Figure 25 representing all the parcels. Figure 27 shows a plot of the relationship between yield and maximum FCover for the planting dates between March 27th to April 14th. Finally, Figure 28 shows the correlation between yield and the maximum FCover for the parcels planted between April 15th and 27th is better than the previously mentioned planting dates, with a correlation value of $R=0.62$.

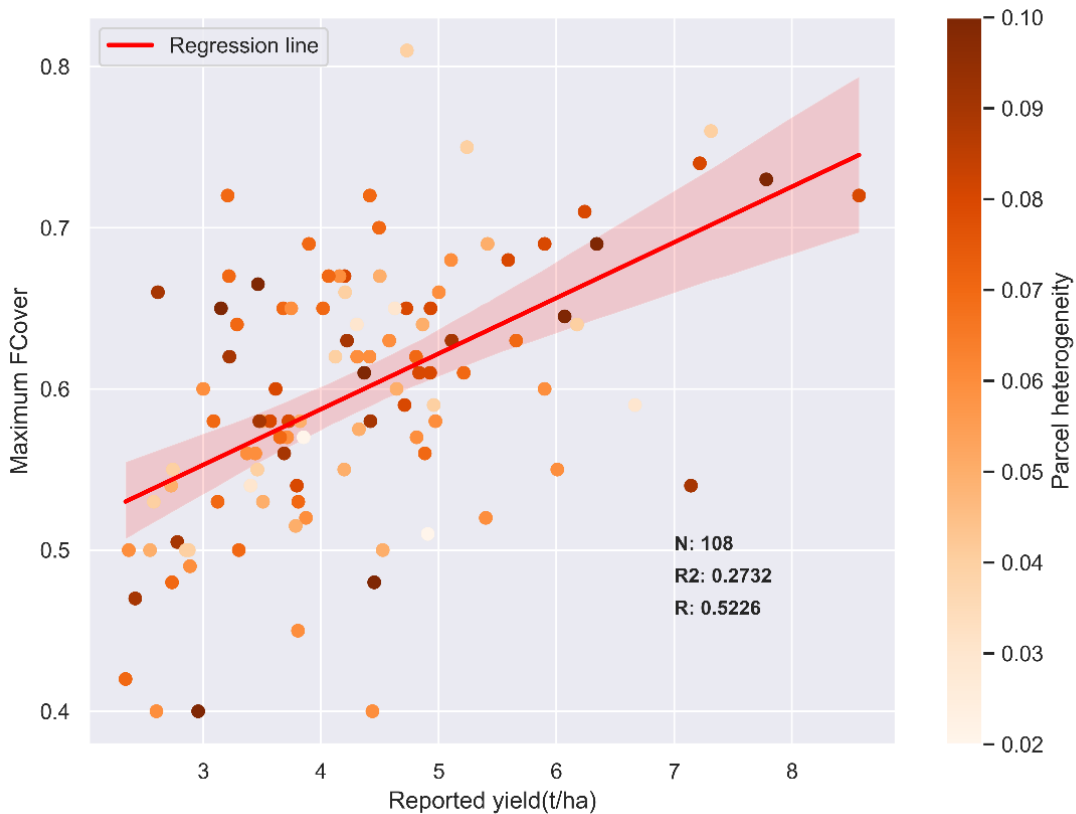


Figure 25: Relationship between the reported yield and maximum FCover for 108 parcels

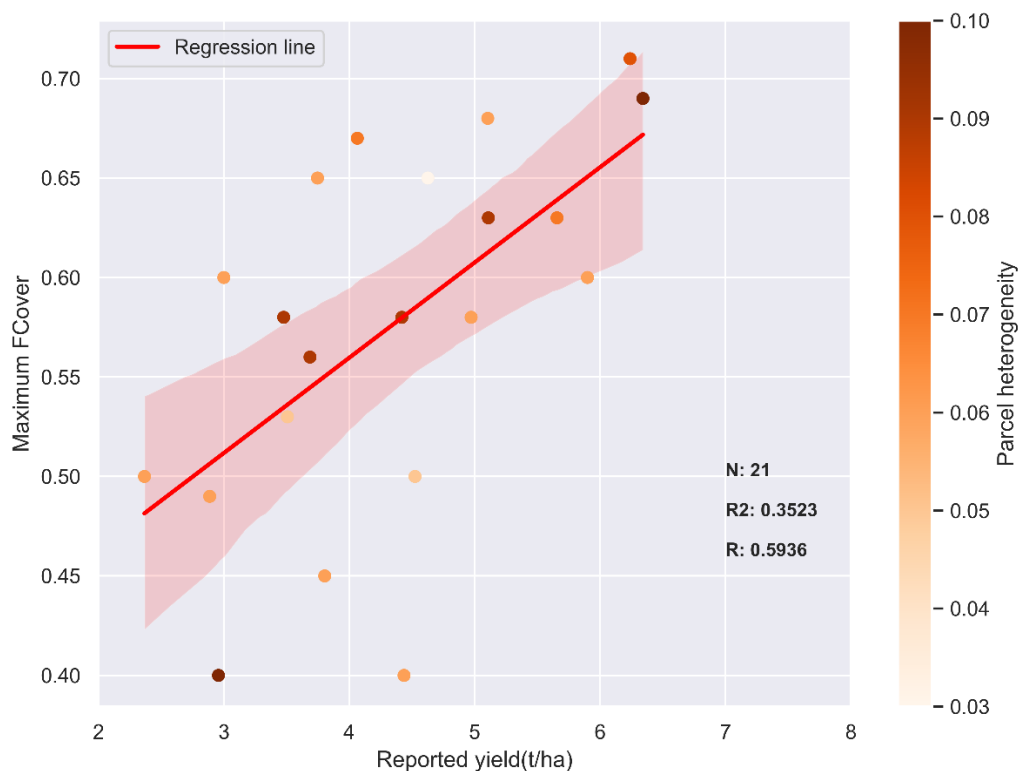


Figure 26: Relationship between the reported yield and maximum FCover for parcels with planting dates between March 3-26

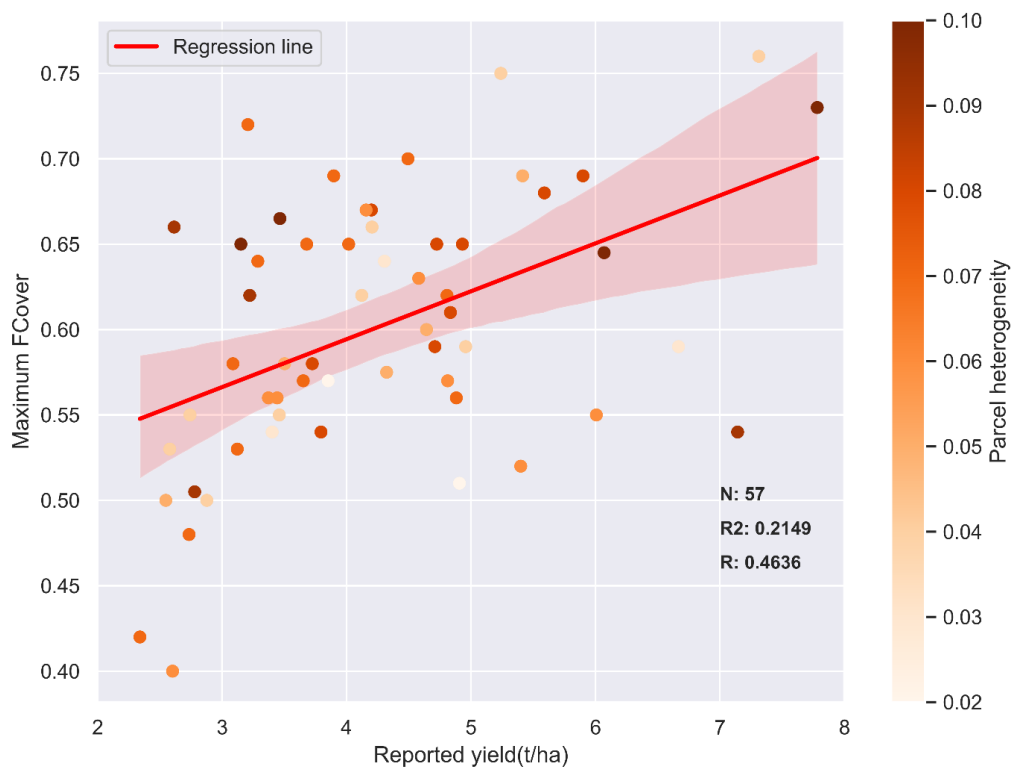


Figure 27: Relationship between the reported yield and maximum FCover for parcels with planting dates between March 27-April 14

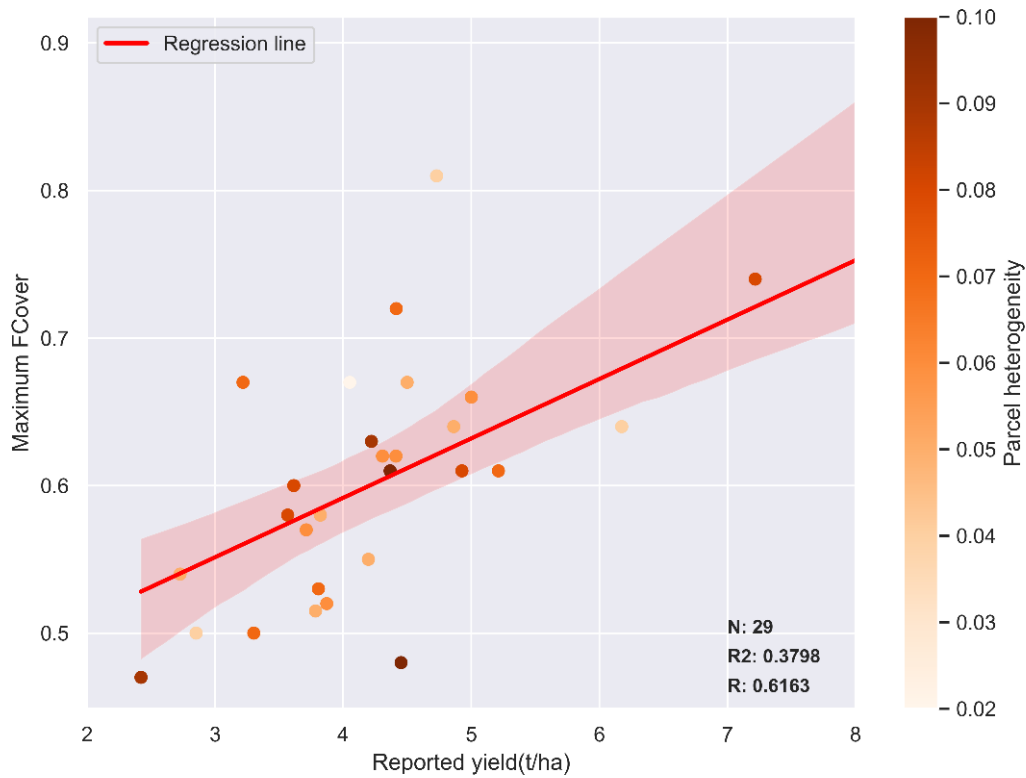


Figure 28: Relationship between the reported yield and maximum FCover for parcels with planting dates between April 15-27, 2021

5.1.5. Sensitivity analysis results

Figure 28 shows the effect of Clay soil on the simulated yield. The yield produced by using Clay soil in all parcels' simulations tends to be within the range of the reported yield. On the other hand, Figure 29 shows the yield simulation obtained using Sandy clay loam for all the parcels, and the model tends to overestimate the yield values in this scenario compared to the reported yield.

The results from the graph in Figure 29 and Figure 30 indicate that the soil parameter should be well-calibrated for the model to estimate the yield correctly. Additionally, the results also suggest that if the soil texture for a particular parcel was wrongly assigned during simulation, the model might produce wrong results. The soil texture dictates the water-holding capacity of the given soil and, as a result, also affects the amount of water available to support a plant's growth.

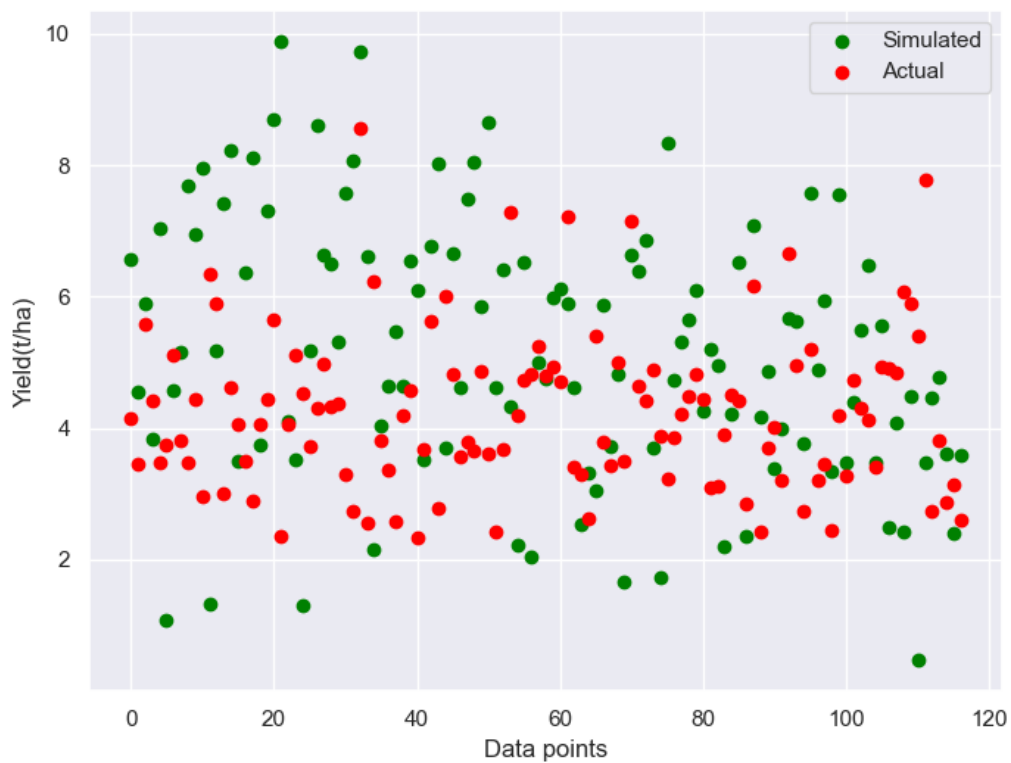


Figure 29: Yield simulation with Clay soil compared with the reported yield

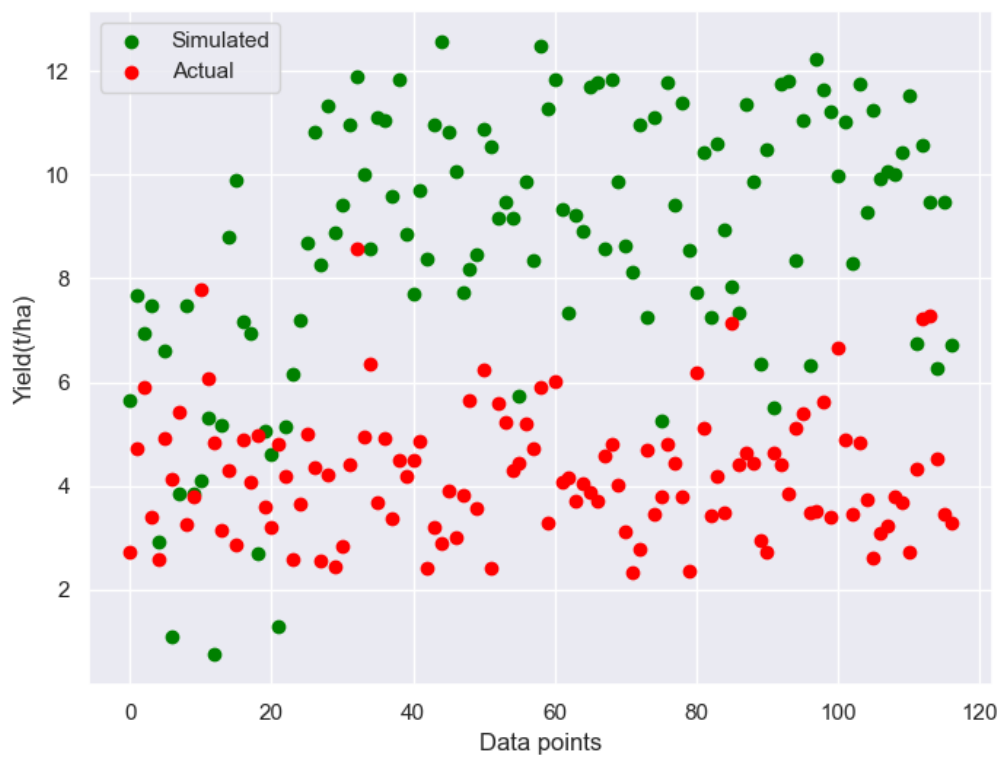


Figure 30: Yield simulation with Sandy clay loam soil compared with reported yield

5.1.6. Cloud masking results

The cloud shadow and layers from FMask were used to mask pixels from the FCover images containing clouds or cloud shadow. FMask managed to mask out a good percentage of clouds within the images, but some residuals of the clouds could still be seen on the images afterwards.

The image before and after the cloud mask processing is shown in Figure 31. The red-outlined polygons indicate some sample areas within the FCover images covered by clouds that were afterwards masked out. After cloud masking, some residual clouds can be seen around the sample areas. Hence, after cloud masking, the resulting FCover images were assessed in QGIS for residual clouds in the parcels' areas to ensure that the cloud mask retrieved the unwanted pixels. Fortunately, for all the selected parcels, no residual clouds were identified within them. Residual cloud cover within the parcel would have resulted in wrong FCover values being included in the analysis; hence it was important to undertake this verification to identify affected pixels that needed to be discarded.

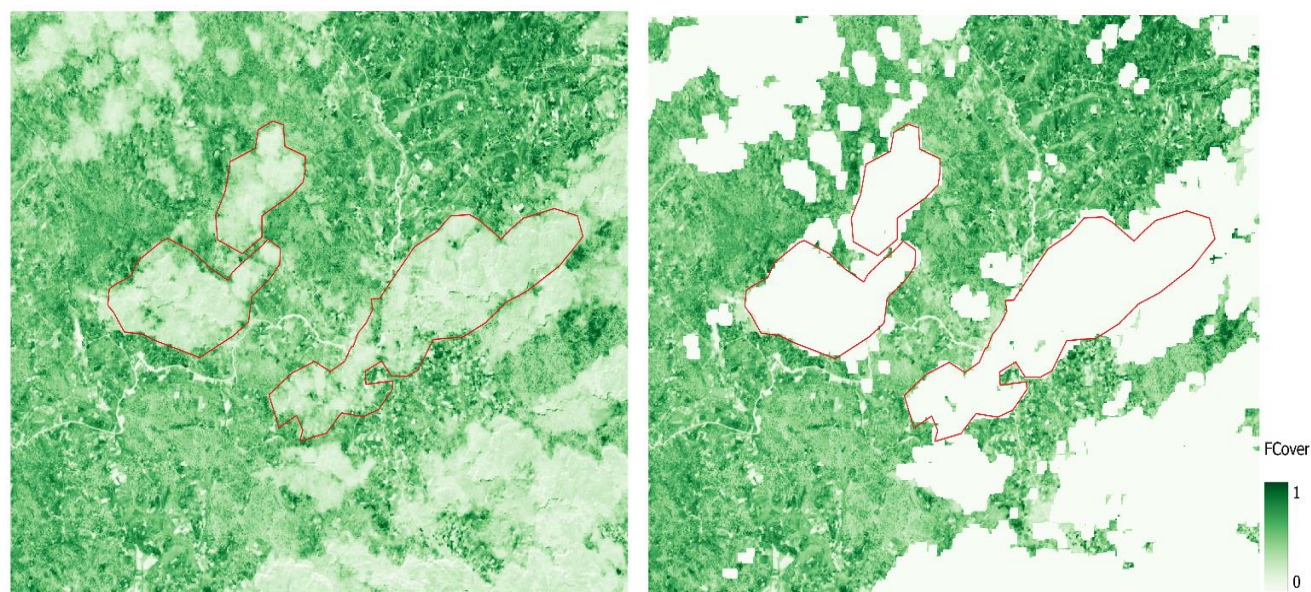


Figure 31: FCover image before (left) and after (right) cloud masking

5.1.7. Yield simulation results before assimilation

The performance estimation of the model, when the simulated yield is compared to the reported yield, is represented in Figure 32. The model has a RMSE of 2.227, R^2 of 0.198, NSE of -2.296

and MAE of 1.876. The level of heterogeneity of the parcels is indicated by the colour intensity, whose scale is represented on the right-hand side of the graph. A negative NSE value indicates a significant systematic error in the model simulated yield results.

The dataset was further divided into three parts to investigate the relationship between the simulated and the reported yield according to the planting dates. The outcome is shown in Figure 33 and (see Appendix 5) Figures 46 and 47. Figure 33 shows a better correlation between the reported and simulated yield for the parcels with planting dates falling between March 3-26 compared to the combination of all the planting dates and planting dates falling between March 26-14 April, and April 15-27.

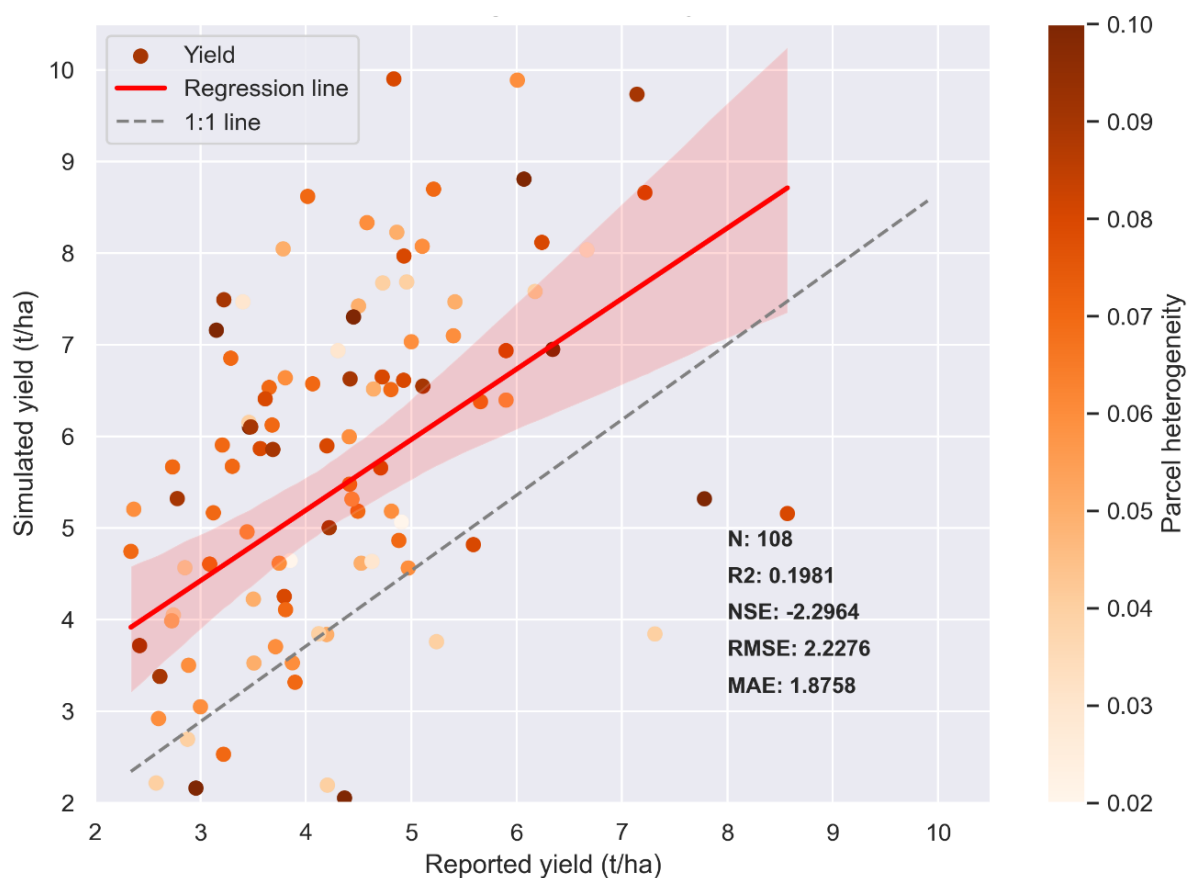


Figure 32: AquaCrop model yield estimation performance before the assimilation of FCover

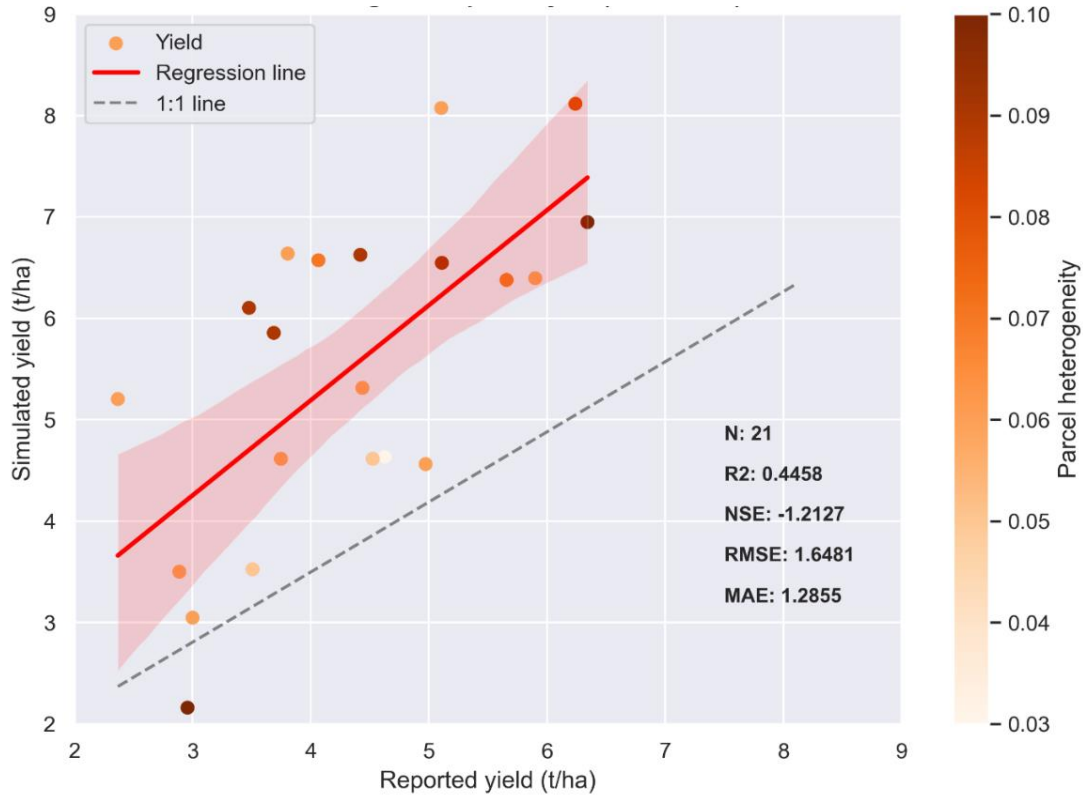


Figure 33: AquaCrop model yield estimation performance before the assimilation for March 3-26 planting dates

The yield estimated through AquaCrop and the reported yield were plotted against other parameters, such as crop cycle duration and the planting dates shown in Figure 34 and Figure 35. The analysis was essential to assess and understand the patterns in the data explaining the effect of soil texture on the crop cycle duration and the simulated yield. In addition, the plot of yield against planting dates helped in understanding the planting date patterns in relation to the soil type.

The yield against crop cycle duration was additionally used to understand the dissimilarity between the simulated and reported yield before FCover's assimilation. Figure 34 shows that, on average, some parcels with Clay soil had a longer crop cycle duration, i.e., greater than 260 days, compared to Sandy clay loam soil parcels.

The crop cycle duration can be explained using the yield distribution map in Figure 8, which shows that areas with Clay soil were located within mid-altitude regions compared to the areas with Sandy clay loam soil, mainly in low-altitude regions. Additionally, Figure 35 shows the reported and simulated yield graph against the planting dates. The graph shows two significant clusters formed

by the data points according to the two planting months, i.e., March and April. The graph indicates that parcels with planting dates in March were mainly located within the Clay soil regions.



Figure 34: Simulated and reported yield against crop cycle duration



Figure 35: Simulated and reported yield against planting dates

Figures 34 and 35 also show a difference between the simulated and reported yield due to the dissimilarity of the points in the two categories, apart from a few parcels with an agreement between the simulated and reported yield values.

5.1.8. Yield simulation results after assimilation

Figure 36 illustrates the model's performance after assimilation, considering the different metrics. The metrics indicate that the model had a significant performance improvement after the assimilation of FCover. The model obtained a RMSE of 1.557, a R^2 value of 0.308, and a MAE value of 1.301. However, the model performance remains relatively low, but there is an improvement of approximately 30% in terms of reduction of RMSE and MAE thanks to the assimilation of FCover. The model's accuracy may be impacted by the parcels' heterogeneity, variety of maize or different management practices within the different parcels. Additionally, the reported yield used in this study is the grain yield which may not be directly represented by information retrieved from the satellite.

Similar to the case before assimilation, the dataset was divided into three according to planting dates to investigate the relationship between the simulated and the reported yield taking the planting dates into account. The outcome is shown in Figure 37 and (see Appendix 6) Figures 48 and 49. Figure 37 shows a better correlation between the reported and simulated yield for the parcels with planting dates falling between March 3 and 26 compared to the combination of all the planting dates and planting dates falling between March 26-14 April and April 15-27.

After the assimilation of FCover in AquaCrop, the model produced a new set of simulated yield results for each parcel. The resulting information was then used to assess the model's performance after assimilation against the reported yield. Figure 38 compares the reported and simulated yield against crop cycle duration. Compared to the case before assimilation in Figure 33, more points representing the reported and the simulated yield tend to agree. Additionally, Figure 39 shows that the relationship between the maximum FCover and the simulated yield has a moderate correlation of $R=0.28$.

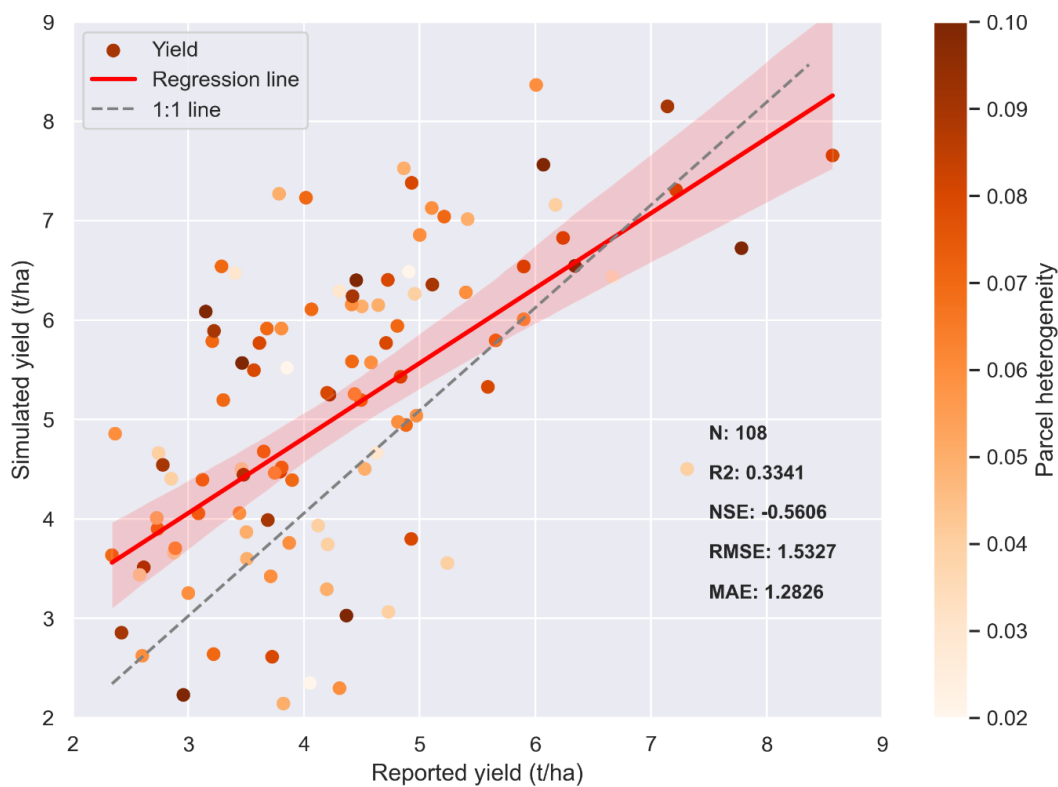


Figure 36: Model yield estimation performance after the assimilation of FCover

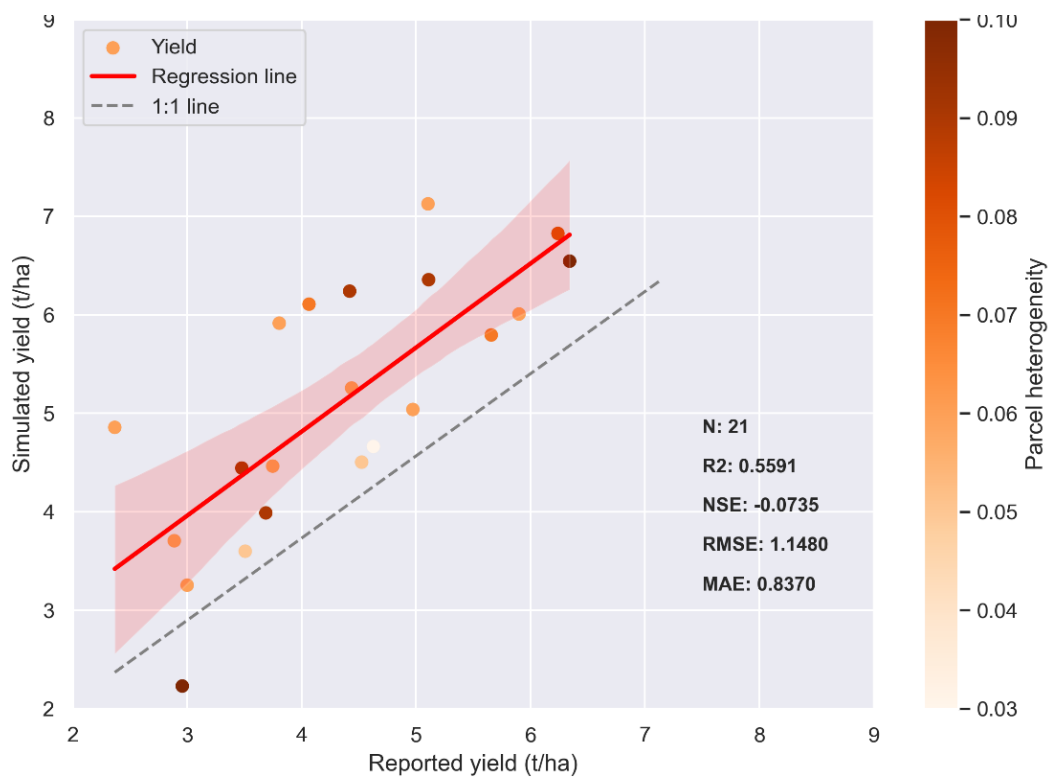


Figure 37: Model yield estimation performance after assimilation for March 3-26 planting dates



Figure 38: Simulated (assimilation) and reported yield against crop cycle duration

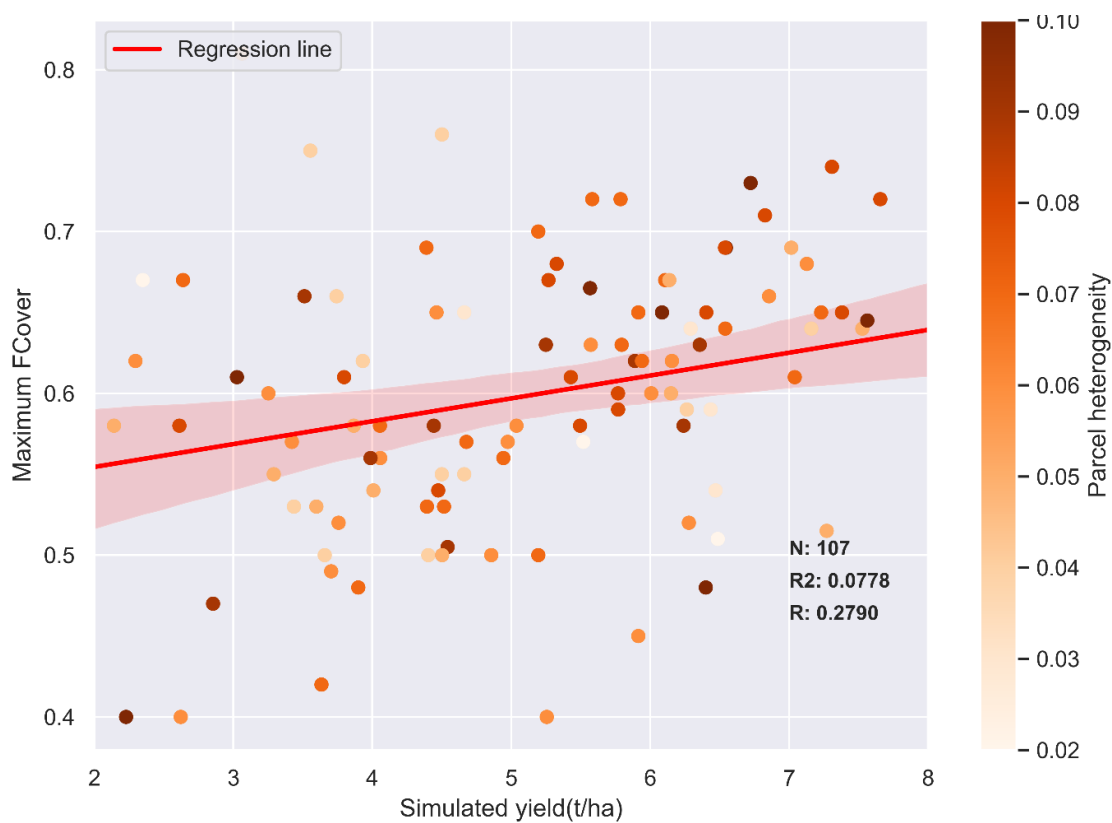


Figure 39: Relationship between simulated yield after assimilation and Maximum FCover

5.1.9. Yield simulation results at ward level

Figures 40 and 41 show the performances achieved before the assimilation of FCover and after assimilation for the ward level after the aggregation of the parcel level yield information. Figure 40 shows that the model achieved a performance of $R^2=0.812$ and $RMSE = 1.552$ before assimilation. The NSE value of -9.856 shows that the model had significant systematic errors when simulating the yield values. Figure 41 shows that after assimilation, there was an improvement in the model's performance with $R^2 = 0.843$ and $RMSE = 0.983$. As expected, the performance of yield estimation at the ward level is much higher than at the parcel level. This improvement is because the random component of the prediction error was smoothened out. Additionally, it is interesting to see that the FCover assimilation also enhances the model's performance after aggregation at the ward level. There is an improvement of approximately 36% in terms of reduction of RMSE and MAE due to the assimilation of FCover.

Table 6 shows the aggregated yield outcome for each ward and the combined wards to obtain 15 parcels for the aggregation exercise.

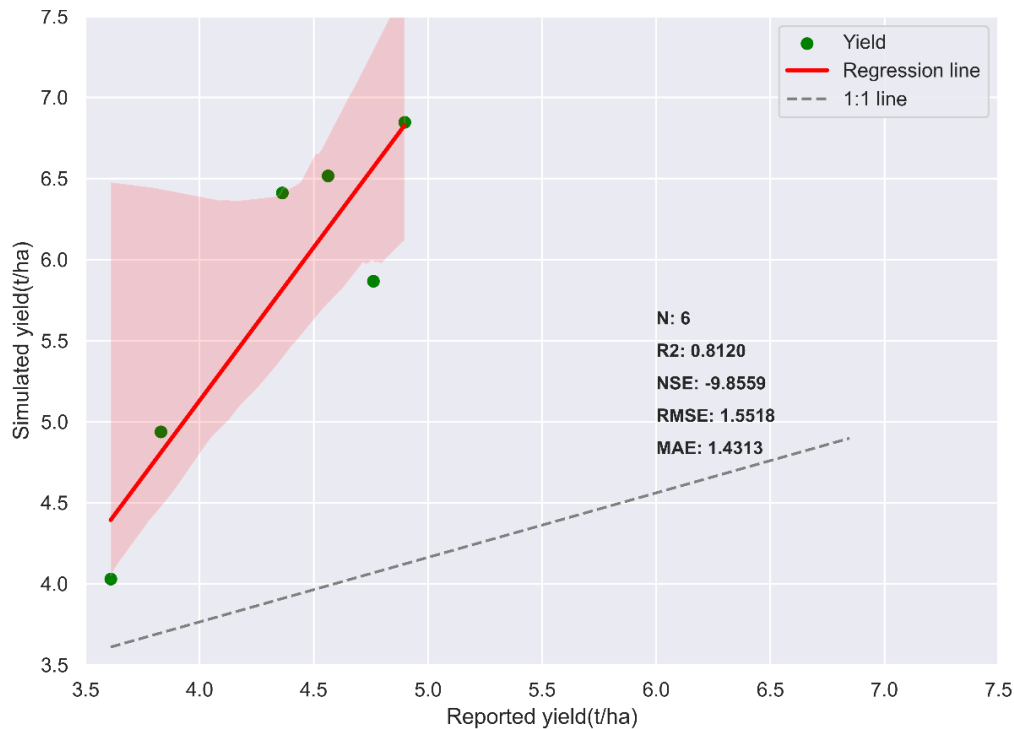


Figure 40: Model yield performance before assimilation aggregated at the ward level

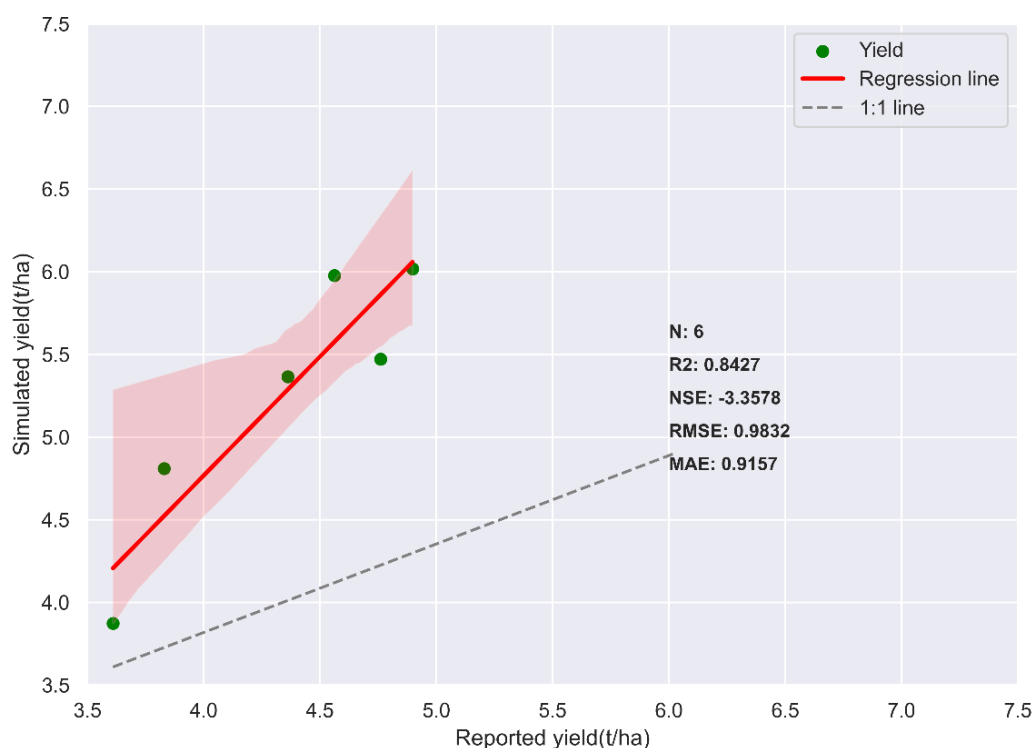


Figure 41: Model yield performance after assimilation aggregated at the ward level

Table 6: Yield estimated at the ward level after aggregation

Ward	Reported Yield	Simulated Yield	Simulated yield after assimilation
Cherang'any/Chebororwa	4.36	6.41	5.36
Kepchemutwa	4.76	5.86	5.47
Karuna	4.89	6.84	6.01
Kamariny	4.56	6.51	5.97
Ngeria	3.61	4.03	3.87
Megun			
Tulwet/chuiyat			
Tarakwa			
Ainakiboi/olare			
Kilombe	3.83	4.94	4.81
Kapsaos			
Kuinet/Kapsuswa			

Table 7 summarises the model's performances at different scales before and after assimilation. Table 8 outlines the comparison of the correlation coefficient using Fisher exact test. The p-value obtained for the parcel and the ward level shows that the difference between the model performance with and without assimilation was not statistically significant. This is because the p-value in both cases is greater than 0.05; hence we reject the null hypothesis.

Table 7: Summary of the model performance

	Before assimilation				After assimilation			
Metrics	R^2	NSE	RMSE	MAE	R^2	NSE	RMSE	MAE
Parcel level	0.198	-8.967	3.873	3.687	0.334	-0.560	1.532	1.282
Ward level	0.812	-9.856	1.552	1.431	0.843	-3.36	0.983	0.916

Table 8: Summary of the comparison of correlation coefficient using the Fisher exact test

		Before assimilation(r_2)	After assimilation(r_1)
Parcel level	r	0.478	0.659
	n	108	108
	z	1.312	
	p-value	0.189	
Ward level	r	1.133	1.231
	n	6	6
	z	0.120	
	p-value	0.904	

5.2. Discussion

5.2.1. Parcel heterogeneity and diversity of crop varieties

The study undertaken by Jin et al. (2019) identifies that most smallholder systems have enormous heterogeneity. In this study, the parcels were investigated for heterogeneity, and it was realized that they were all heterogeneous to some degree. Parcels with a high heterogeneity were investigated, and the findings were that the difference in FCover values within a parcel was sometimes caused by the presence of a tree growing within the parcel or uneven vegetation density. The point within a parcel where trees grow may register higher FCover values than the areas with growing maize crops. When using satellite data, this may cause the wrong indication of a high maximum FCover value within the parcel.

Aside from the issue of heterogeneity, the region under investigation was characterised by farmers planting their crops during different months and dates. This brought out the great difference in planting dates in small-holder systems, as Lobell et al. (2020) argued. Additionally, the difference in planting dates also influenced the duration of the crop growth cycle in that the crops planted earlier had a longer crop cycle duration than those planted later, which may be because farmers may decide to grow different varieties of maize during other planting dates based on the varieties ability to withstand certain conditions as pointed out by Gachara et al. (2022). Different maize varieties have different growth capabilities, such as canopy expansion, that may sometimes not directly relate to the grain yield produced. Altitude also affected the crop cycle duration since the crops grown at lower altitudes took a shorter crop cycle duration than their counterparts grown in higher altitudes, which can be attributed to the effect of temperature on plant growth.

5.2.2. Model Sensitivity Analysis

The model sensitivity analysis demonstrated that soil is one of the parameters that should be well defined in the model as it affected, to a large extent, the outcome of the yield values obtained. Clay soil produced a lower yield than Sandy clay loam soil.

The difference in the yield outcome between Clay and Sandy clay loam may be because soil texture influences the soil's water-holding capacity, affecting water availability to crops. Specifically, the ability of soil to retain and transfer water to crops is impacted by soil texture. Therefore, accurate

measurement and input of soil texture into the Aquacrop model is crucial to obtaining reliable simulation results.

The AquaCrop model provides reference soil parameters for all types of soil textures, which reduces the difficulty of field soil measurements. However, the provided reference soil parameters may not always be able to accurately simulate the exact soil characteristics within the parcels; this may be due to variables such as climatic conditions and field management practices that are difficult to account for in the simulations. Lobell et al. (2020) argue that the quality of satellite estimates such as FCover may also affect the output of a model. However, most of the FCover profiles obtained in this study were quite smooth, supporting the idea of efficient and robust retrieval of FCover from Sentinel-2 data.

5.2.3. FCover Time Series

According to Delrue et al. (2013), the presence of cloud cover in many small-scale agricultural areas makes it challenging to obtain a complete time series covering all the stages of the crop-growing cycle. The presence of cloud cover was one of the challenges experienced when processing the FCover time series of parcels as Kenya is located along the equator and hence prone to the problem of cloud cover, especially during the planting season. Nevertheless, cloud cover did not severely impact the data obtained from remote sensing in this study. The Savitzky Golay method was employed to undertake gap-filling and smoothing to obtain a smooth profile covering the whole time series where the satellite images were affected by cloud cover and to fill the gap between the satellite acquisition dates.

The FCover time series gave a good indication of the different phenological stages of the crop growth within the parcel, such as emergence and senescence, which made it possible to obtain the emergence date from it, according to the procedure used by Mohamed Sallah et al. (2019).

5.2.4. Yield estimation performance

AquaCrop performance achieved in this study at the ward level before assimilation was $RMSE = 1.55$ and $R^2 = 0.812$. After assimilation, the performance improved to $RMSE=0.98$ and $R^2 = 0.84$. The performance achieved in this study compared to the results obtained by Todorovic et al. (2009)

of RMSE = 0.70 without assimilation of biophysical variables can be considered slightly low. Todorovic et al. (2009) achieved a better accuracy in estimating yield using Aquacrop since the model was calibrated, while in this study, the model's default parameters were used. After assimilation, the model performance of $R^2 = 0.84$ at the ward level, when compared to the results obtained at a regional level by Jin et al. (2020) of $R^2 = 0.78$ and by Abi Saab et al. (2021) of $R^2 = 0.54$ this study obtained a better performance. In general, the model does not require the calibration of parameters to obtain yield estimates, but from other studies that employed this additional step, the model had a better performance, such as the study undertaken by Todorovic et al. (2009).

The studies undertaken by Todorovic et al. (2009), Jin et al. (2020) and Abi Saab et al. (2021) also entailed field experiments which included soil tests being undertaken to ensure that the parameters used in the model were well established. Field experiments are also important to ensure that the researcher is aware of the quality of the dataset collected. In the case of this study, field experiments were not possible as the dataset obtained from Bird's-Eye was acquired in 2021. In addition, as much as they were regarded as homogeneous, the parcels selected for this study still possessed a degree of heterogeneity.

The difference in AquaCrop's performance in yield estimation from this study and the other previous studies may be due to the reasons Roell et al. (2020) brought out. Roell et al. (2020) argue that data quality and quantity differences can significantly impact the model's performance. In addition, different geographic regions have different climatic conditions, and farmers may opt for different varieties of the same crop, as in the case of Western Kenya brought out by Gachara et al. (2022), which may challenge the model's calibration. Furthermore, the level of estimation may result in different model performance outcomes. As in the case of this study, before assimilation at the parcel level, the performance achieved was $R^2 = 0.334$ and RMSE =1.532, while at the ward level $R^2=0.843$ and RMSE =0.983. The outcome shows that the model performs better at the ward than at the parcel level since the random component of the prediction error was smoothened out after aggregation. Unfortunately, the in situ yield measurement limits the number of ward level estimation preventing to conclude in a statistically sound manner.

Furthermore, the NSE values obtained for comparing the simulated and reported yields indicate that the model had a significant systematic error in simulating yield. Before the assimilation of FCover, the NSE and RMSE errors seemed to be high but reduced once FCover had been assimilated into the model. This outcome indicates that assimilation also reduces the systematic errors in a model when simulating the yield.

The comparison of R^2 and RMSE metrics before and after assimilation indicates that the model performance improved with the assimilation of FCover. However, the statistical test undertaken using Fisher exact test suggests that the difference is not statistically significant. This result obtained using the Fisher test implies that the assimilation of FCover may have improved the model's performance in this study but did not have a significant impact.

CHAPTER 6. Conclusion and Recommendations

This chapter summarises our findings from this study, including the factors that led to the success or failure of the outcome of this study. Additionally, it provides recommendations on what can be improved for further studies in this research area.

6.1. Conclusion

In conclusion, this study's main objective was to establish if remotely sensed biophysical variables can improve the performance of AquaCrop in crop yield estimation. The results obtained in this study for both parcel and ward levels have proven that when remote sensing data is coupled with a crop yield model, the model's performance improves. However, due to the lack of calibration of the model in this study, the model had significant systematic errors that render the model not very reliable unless the calibration of some conservative model parameters is undertaken. Generally, when coupled with the remote sensing dataset, the model improves yield estimation. However, we have to consider that remote sensing is not a perfect solution due to its limitations. When undertaken precisely and using reliable in situ datasets, the model may serve as a decision-making tool for policymakers in a given country such as Kenya.

6.2. Recommendations

Based on the results of this study, some recommendations can be made to improve maize crop yield estimation using the AquaCrop model. These recommendations aim to enhance the accuracy of the modelling outcome for better decision-making.

The first recommendation entails the use of the models' default parameters. Previous studies, such as the one undertaken by Todorovic et al. (2009), showed that the model performed better than the outcome in this study when some conservative parameters were calibrated. Calibration is useful in reducing the model's systematic errors, hence ensuring reliable results.

The in situ dataset used in this study contained helpful information that assisted in the modelling process—however, some of the missing information made obtaining accurate estimations with the model challenging. For example, the initial water content before planting the crop obtained through field measurements would have provided reliable values.

The soil texture information may be verified through field soil tests to ensure that the soil textures are classified precisely by the percentage of the different soil compositions.

Additionally, it is important to verify the parcels to check if features such as trees may interfere with the FCover values providing a false-positive indication of high yield through the information obtained from satellite images.

In preprocessing the satellite time series, using FMask to mask out clouds resulted in cloud residuals within the image that may result in invalid FCover time series data. This challenge may be overcome by using a different cloud mask with better performance or by decreasing the threshold of FMask to a value lower than 10%.

The assimilation method used in this study should be compared against other assimilation techniques to understand how the assimilation technique also affects the outcome.

Finally, the reported field dataset used should be of good quality that can be used in the validation of the simulated dataset and the calibration of the model.

CHAPTER 7. References

- Abedinpour, M., Sarangi, A., Rajput, T.B.S., Singh, M., Pathak, H., Ahmad, T., 2012. Performance evaluation of AquaCrop model for maize crop in a semi-arid environment. *Agric. Water Manag.* 110, 55–66. <https://doi.org/10.1016/j.agwat.2012.04.001>
- Abi Saab, M.T., El Alam, R., Jomaa, I., Skaf, S., Fahed, S., Albrizio, R., Todorovic, M., 2021. Coupling Remote Sensing Data and AquaCrop Model for Simulation of Winter Wheat Growth under Rainfed and Irrigated Conditions in a Mediterranean Environment. *Agronomy* 11, 2265. <https://doi.org/10.3390/agronomy11112265>
- Ahmadi, S.H., Reis Ghorra, M.R., Sepaskhah, A.R., 2022. Parameterizing the AquaCrop model for potato growth modeling in a semi-arid region. *Field Crops Res.* 288, 108680. <https://doi.org/10.1016/j.fcr.2022.108680>
- Anaconda, n.d. Anaconda. URL <https://anaconda.org/anaconda/jupyter> (accessed 8.2.22).
- Araya, A., Habtu, S., Hadgu, K.M., Kebede, A., Dejene, T., 2010a. Test of AquaCrop model in simulating biomass and yield of water deficient and irrigated barley (*Hordeum vulgare*). *Agric. Water Manag.* 97, 1838–1846. <https://doi.org/10.1016/j.agwat.2010.06.021>
- Araya, A., Keesstra, S.D., Stroosnijder, L., 2010b. Simulating yield response to water of Teff (*Eragrostis tef*) with FAO’s AquaCrop model. *Field Crops Res.* 116, 196–204. <https://doi.org/10.1016/j.fcr.2009.12.010>
- Basso, B., Liu, L., 2019. Seasonal crop yield forecast: Methods, applications, and accuracies, in: *Advances in Agronomy*. Elsevier, pp. 201–255. <https://doi.org/10.1016/bs.agron.2018.11.002>
- Bello, Z.A., Walker, S., 2016. Calibration and validation of AquaCrop for pearl millet (*Pennisetum glaucum*). *Crop Pasture Sci.* 67, 948. <https://doi.org/10.1071/CP15226>
- Chivasa, W., Mutanga, O., Biradar, C., 2017. Application of remote sensing in estimating maize grain yield in heterogeneous African agricultural landscapes: a review. *Int. J. Remote Sens.* 38, 6816–6845. <https://doi.org/10.1080/01431161.2017.1365390>
- Dawson, T.P., Perryman, A.H., Osborne, T.M., 2016. Modelling impacts of climate change on global food security. *Clim. Change* 134, 429–440. <https://doi.org/10.1007/s10584-014-1277-y>

- Delécolle, R., Maas, S.J., Guérif, M., Baret, F., 1992. Remote sensing and crop production models: present trends. *ISPRS J. Photogramm. Remote Sens.* 47, 145–161.
[https://doi.org/10.1016/0924-2716\(92\)90030-D](https://doi.org/10.1016/0924-2716(92)90030-D)
- Delloye, C., Weiss, M., Defourny, P., 2018. Retrieval of the canopy chlorophyll content from Sentinel-2 spectral bands to estimate nitrogen uptake in intensive winter wheat cropping systems. *Remote Sens. Environ.* 216, 245–261. <https://doi.org/10.1016/j.rse.2018.06.037>
- Delrue, J., Bydekerke, L., Eerens, H., Gilliams, S., Piccard, I., Swinnen, E., 2013. Crop mapping in countries with small-scale farming: a case study for West Shewa, Ethiopia. *Int. J. Remote Sens.* 34, 2566–2582. <https://doi.org/10.1080/01431161.2012.747016>
- Dhakar, R., Sehgal, V.K., Chakraborty, D., Sahoo, R.N., Mukherjee, J., Ines, A.V.M., Kumar, S.N., Shirsath, P.B., Roy, S.B., 2022. Field scale spatial wheat yield forecasting system under limited field data availability by integrating crop simulation model with weather forecast and satellite remote sensing. *Agric. Syst.* 195, 103299.
<https://doi.org/10.1016/j.agry.2021.103299>
- ECMWF, 2018. ERA-5 land. URL <https://cds.climate.copernicus.eu> (accessed 11.9.22).
- ESA, 2014. Copernicus open access hub. URL <https://scihub.copernicus.eu/> (accessed 11.17.22).
- European Commission. Joint Research Centre. Institute for Environment and Sustainability., CCAFS., 2012. Combining crop models and remote sensing for yield prediction - concepts, applications and challenges for heterogeneous, smallholder environments: report of CCFAS JRC Workshop at Joint Research Centre, Ispra, Italy, June 13 14, 2012. Publications Office, LU.
- European Space Agency, 2023. COPERNICUS. URL <https://spacedata.copernicus.eu/en/web/guest/collections/copernicus-digital-elevation-model> (accessed 3.10.23).
- FAO, 2023. ETo Calculator. URL <https://www.fao.org/land-water/databases-and-software/eto-calculator/en/> (accessed 5.2.23).
- FAO, 2022. AquaCrop Reference Manual, Chapter 3 (No. Version 7.0).
- FAO, 2013. Harmonized World Soil Database. URL <https://gaez.fao.org/pages/hwsd> (accessed 11.22.22).
- FAO, 2009. Evapotranspiration from a reference surface (Reference Manual No. Version 3.1).

- FAO, n.d. AquaCrop-OSPy. URL
https://aquacropos.github.io/aquacrop/notebooks/AquaCrop_OSPy_Notebook_3/
- Fisher, R.A., 1921. On the " Probable Error" of a Coefficient of Correlation Deduced from a Small Sample.
- Foster, T., Brozović, N., Butler, A.P., Neale, C.M.U., Raes, D., Steduto, P., Fereres, E., Hsiao, T.C., 2017. AquaCrop-OS: An open source version of FAO's crop water productivity model. *Agric. Water Manag.* 181, 18–22. <https://doi.org/10.1016/j.agwat.2016.11.015>
- Gachara, G., Suleiman, R., El Kadili, S., Ait Barka, E., Kilima, B., Lahlali, R., 2022. Drivers of Post-Harvest Aflatoxin Contamination: Evidence Gathered from Knowledge Disparities and Field Surveys of Maize Farmers in the Rift Valley Region of Kenya. *Toxins* 14, 618. <https://doi.org/10.3390/toxins14090618>
- Gupta, H.V., Kling, H., 2011. On typical range, sensitivity, and normalization of Mean Squared Error and Nash-Sutcliffe Efficiency type metrics: TECHNICAL NOTE. *Water Resour. Res.* 47. <https://doi.org/10.1029/2011WR010962>
- Hayes, M.J., Decker, W.L., 1996. Using NOAA AVHRR data to estimate maize production in the United States Corn Belt. *Int. J. Remote Sens.* 17, 3189–3200. <https://doi.org/10.1080/01431169608949138>
- Hsiao, T.C., Heng, L., Steduto, P., Rojas-Lara, B., Raes, D., Fereres, E., 2009. AquaCrop—The FAO Crop Model to Simulate Yield Response to Water: III. Parameterization and Testing for Maize. *Agron. J.* 101, 448–459. <https://doi.org/10.2134/agronj2008.0218s>
- Jin, X., Li, Z., Feng, H., Ren, Z., Li, S., 2020. Estimation of maize yield by assimilating biomass and canopy cover derived from hyperspectral data into the AquaCrop model. *Agric. Water Manag.* 227, 105846. <https://doi.org/10.1016/j.agwat.2019.105846>
- Jin, Z., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M., Lobell, D.B., 2019. Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sens. Environ.* 228, 115–128. <https://doi.org/10.1016/j.rse.2019.04.016>
- Kenya Agricultural Research Laboratories-Kenya, Wamari, J., Isaya, S., Kenya Agricultural Research Laboratories-Kenya, Kheng, L., Soil and Water management and Crop Nutrition Section, Miriti, J., National Agricultural Research Centre -Kenya, Obutiati, E., Kenya Agricultural Research institute, 2007. Use of Aquacrop Model to Predict Maize

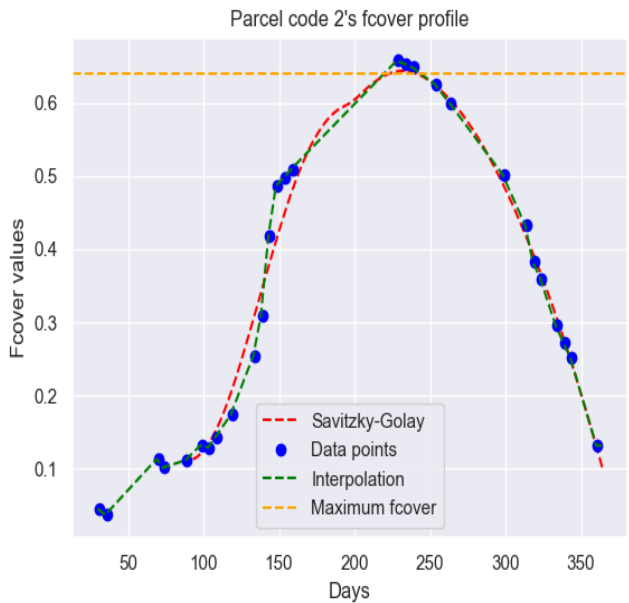
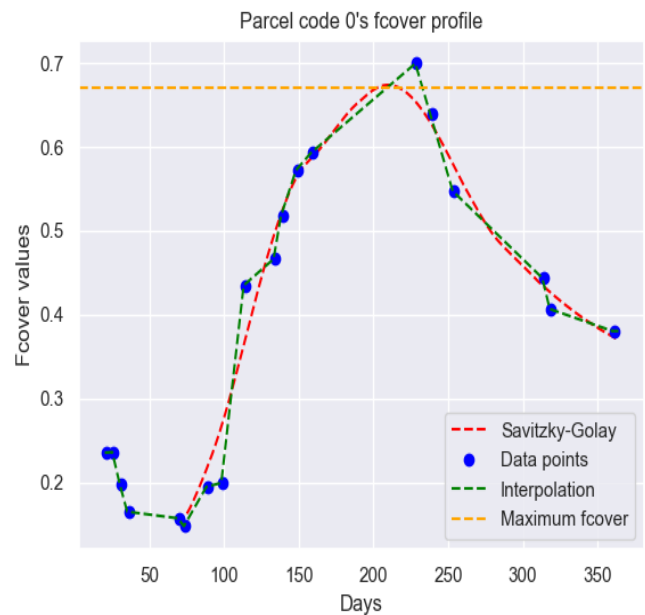
- Yields under varying Rainfall and Temperature in a Semi-Arid Environment in Kenya. J. Meteorol. Relat. Sci. 6, 26–35. <https://doi.org/10.20987/jmrs.08.2012.603>
- Kenya County Guide, 2023. Kenya County Guide. URL kenyacountyguide.co.ke/elgeyo-marakwet-county/ (accessed 5.28.23).
- Kenya National Bureau of Statistics, 2019. 2019 Kenya Population and Housing Census (No. Volume I).
- Kenya Soil Survey, 2018. Soil and Terrain Database for Kenya. URL <http://capsella.madgik.di.uoa.gr/dataset/soil-and-terrain-database-for-kenya-kensoter-v2-0> (accessed 3.10.23).
- Lindblom, J., Lundström, C., Ljung, M., Jonsson, A., 2017. Promoting sustainable intensification in precision agriculture: review of decision support systems development and strategies. *Precis. Agric.* 18, 309–331. <https://doi.org/10.1007/s11119-016-9491-4>
- Lobell, D.B., Schlenker, W., Costa-Roberts, J., 2011. Climate Trends and Global Crop Production Since 1980. *Science* 333, 616–620. <https://doi.org/10.1126/science.1204531>
- Lyle, G., Lewis, M., Ostendorf, B., 2013. Testing the Temporal Ability of Landsat Imagery and Precision Agriculture Technology to Provide High Resolution Historical Estimates of Wheat Yield at the Farm Scale. *Remote Sens.* 5, 1549–1567. <https://doi.org/10.3390/rs5041549>
- Mandal, D., Rao, Y.S., 2020. SASYA: An integrated framework for crop biophysical parameter retrieval and within-season crop yield prediction with SAR remote sensing data. *Remote Sens. Appl. Soc. Environ.* 20, 100366. <https://doi.org/10.1016/j.rsase.2020.100366>
- Mcharo, M., Maghenda, M., 2021. Cost-benefit analysis of sustainable land and water management practices in selected highland water catchments of Kenya. *Sci. Afr.* 12, e00779. <https://doi.org/10.1016/j.sciaf.2021.e00779>
- Mibulo, T., Kiggundu, N., 2018. Evaluation of FAO AquaCrop Model for Simulating Rainfed Maize Growth and Yields in Uganda. *Agronomy* 8, 238. <https://doi.org/10.3390/agronomy8110238>
- Mitran, T., Meena, R.S., Chakraborty, A. (Eds.), 2021. *Geospatial Technologies for Crops and Soils*. Springer Singapore, Singapore. <https://doi.org/10.1007/978-981-15-6864-0>
- Mohamed Sallah, A.-H., Tychon, B., Piccard, I., Gobin, A., Van Hoolst, R., Djaby, B., Wellens, J., 2019. Batch-processing of AquaCrop plug-in for rainfed maize using satellite derived

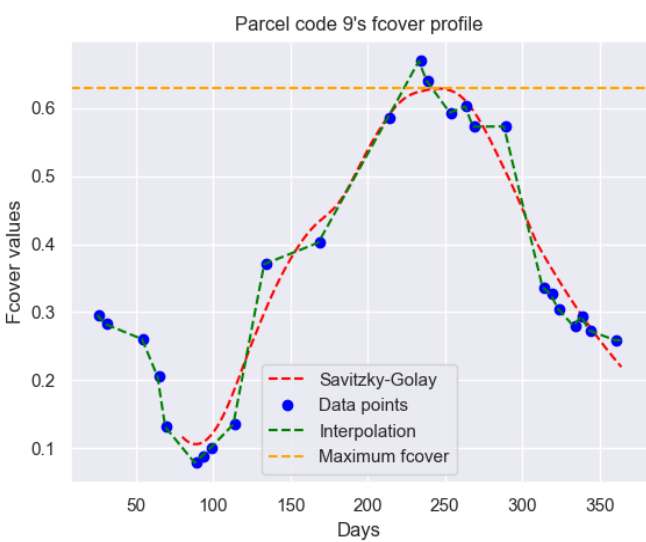
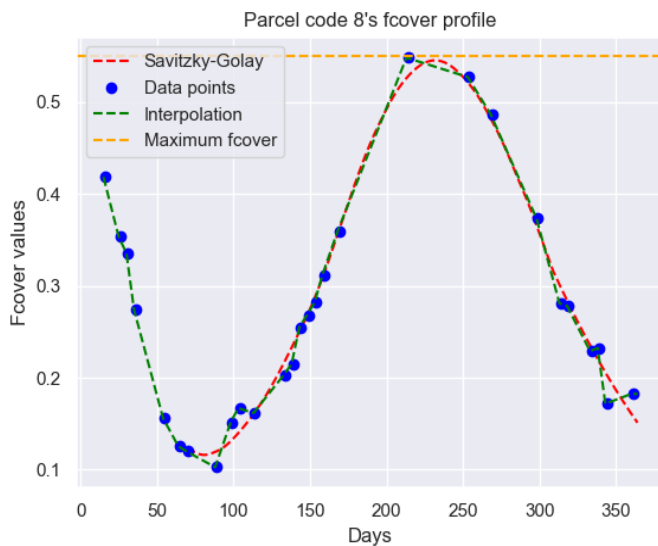
- Fractional Vegetation Cover data. *Agric. Water Manag.* 217, 346–355.
<https://doi.org/10.1016/j.agwat.2019.03.016>
- Niedźwiecki, M.J., Ciołek, M., Gańcza, A., Kaczmarek, P., 2021. Application of regularized Savitzky–Golay filters to identification of time-varying systems. *Automatica* 133, 109865. <https://doi.org/10.1016/j.automatica.2021.109865>
- Ozdogan, M., Yang, Y., Allez, G., Cervantes, C., 2010. Remote Sensing of Irrigated Agriculture: Opportunities and Challenges. *Remote Sens.* 2, 2274–2304.
<https://doi.org/10.3390/rs2092274>
- Pierce, M., 2016. Sustainable Household Food Security in Moiben, Uasin Gishu County, Kenya. *Food Stud. Interdiscip. J.* 6, 77.
- Porwollik, V., Müller, C., Elliott, J., Chrysanthacopoulos, J., Iizumi, T., Ray, D.K., Ruane, A.C., Arneth, A., Balkovič, J., Ciais, P., Deryng, D., Folberth, C., Izaurrealde, R.C., Jones, C.D., Khabarov, N., Lawrence, P.J., Liu, W., Pugh, T.A.M., Reddy, A., Sakurai, G., Schmid, E., Wang, X., De Wit, A., Wu, X., 2017. Spatial and temporal uncertainty of crop yield aggregations. *Eur. J. Agron.* 88, 10–21.
<https://doi.org/10.1016/j.eja.2016.08.006>
- PULA, 2022. Bird’s-Eye. URL <https://ecaas.pula.io/> (accessed 11.2.22).
- Roell, Y.E., Beucher, A., Møller, P.G., Greve, M.B., Greve, M.H., 2020. Comparing a Random Forest Based Prediction of Winter Wheat Yield to Historical Yield Potential. *Agronomy* 10, 395. <https://doi.org/10.3390/agronomy10030395>
- Sherman, G., 2002. QGIS. URL <https://www.qgis.org/en/site/forusers/download.html> (accessed 5.3.20).
- Szostak, M., Pietrzykowski, M., Likus-Cieślak, J., 2020. Reclaimed Area Land Cover Mapping Using Sentinel-2 Imagery and LiDAR Point Clouds. *Remote Sens.* 12, 261.
<https://doi.org/10.3390/rs12020261>
- Todorovic, M., Albrizio, R., Zivotic, L., Saab, M.A., Stöckle, C., Steduto, P., 2009. Assessment of AquaCrop, CropSyst, and WOFOST Models in the Simulation of Sunflower Growth under Different Water Regimes. *Agron. J.* 101, 509–521.
<https://doi.org/10.2134/agronj2008.0166s>
- U.S. Department of Agriculture, 2017. Soil Survey Manual (No. Soil Conservation Service Volume Handbook 18), Soil Survey Division Staff.

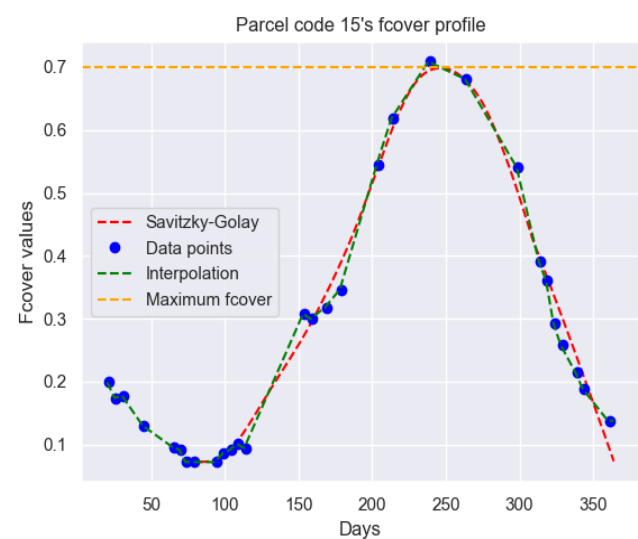
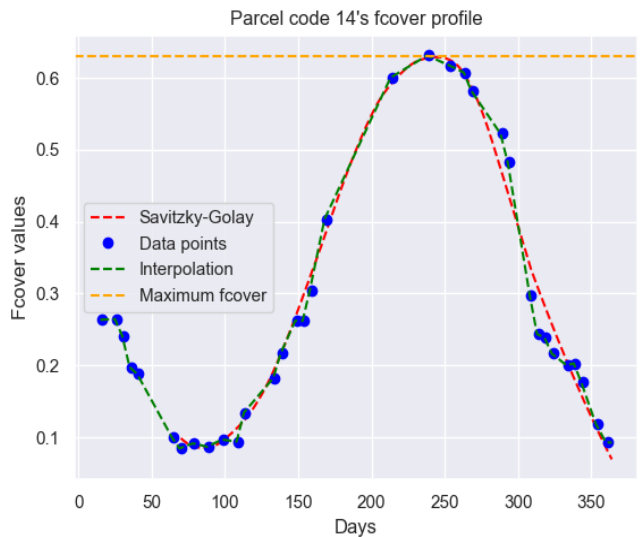
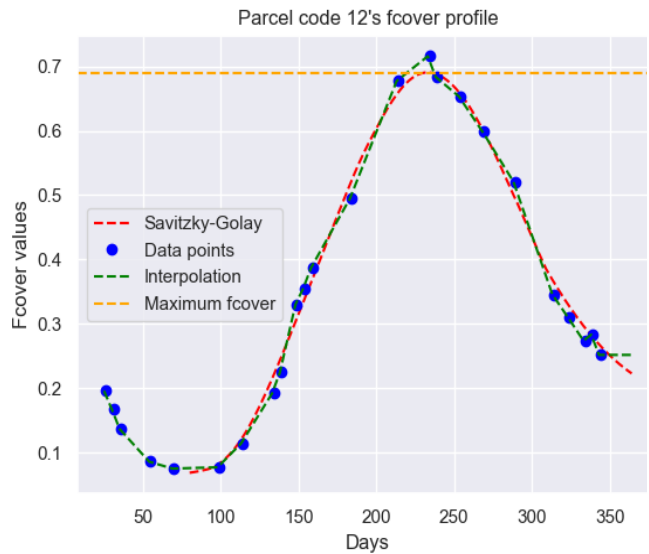
- Vanuytrecht, E., Raes, D., Steduto, P., Hsiao, T.C., Fereres, E., Heng, L.K., Garcia Vila, M., Mejias Moreno, P., 2014. AquaCrop: FAO's crop water productivity and yield response model. *Environ. Model. Softw.* 62, 351–360.
<https://doi.org/10.1016/j.envsoft.2014.08.005>
- Viana, C.M., Freire, D., Abrantes, P., Rocha, J., Pereira, P., 2022. Agricultural land systems importance for supporting food security and sustainable development goals: A systematic review. *Sci. Total Environ.* 806, 150718. <https://doi.org/10.1016/j.scitotenv.2021.150718>
- Waldner, F., Lambert, M.-J., Li, W., Weiss, M., Demarez, V., Morin, D., Marais-Sicre, C., Hagolle, O., Baret, F., Defourny, P., 2015. Land Cover and Crop Type Classification along the Season Based on Biophysical Variables Retrieved from Multi-Sensor High-Resolution Time Series. *Remote Sens.* 7, 10400–10424.
<https://doi.org/10.3390/rs70810400>
- Weiss, M., Baret, F., 1999. Evaluation of Canopy Biophysical Variable Retrieval Performances from the Accumulation of Large Swath Satellite Data. *Remote Sens. Environ.* 70, 293–306. [https://doi.org/10.1016/S0034-4257\(99\)00045-0](https://doi.org/10.1016/S0034-4257(99)00045-0)
- Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* 236, 111402. <https://doi.org/10.1016/j.rse.2019.111402>
- Wellens, J., Raes, D., Fereres, E., Diels, J., Copppe, C., Adiele, J.G., Ezui, K.S.G., Becerra, L.-A., Selvaraj, M.G., Dercon, G., Heng, L.K., 2022. Calibration and validation of the FAO AquaCrop water productivity model for cassava (*Manihot esculenta* Crantz). *Agric. Water Manag.* 263, 107491. <https://doi.org/10.1016/j.agwat.2022.107491>
- Wellens, J., Raes, D., Traore, F., Denis, A., Djaby, B., Tychon, B., 2013. Performance assessment of the FAO AquaCrop model for irrigated cabbage on farmer plots in a semi-arid environment. *Agric. Water Manag.* 127, 40–47.
<https://doi.org/10.1016/j.agwat.2013.05.012>
- World Food Programme, 2022. Impact of increasing fertilizer prices on maize production in Kenya, September 2022.

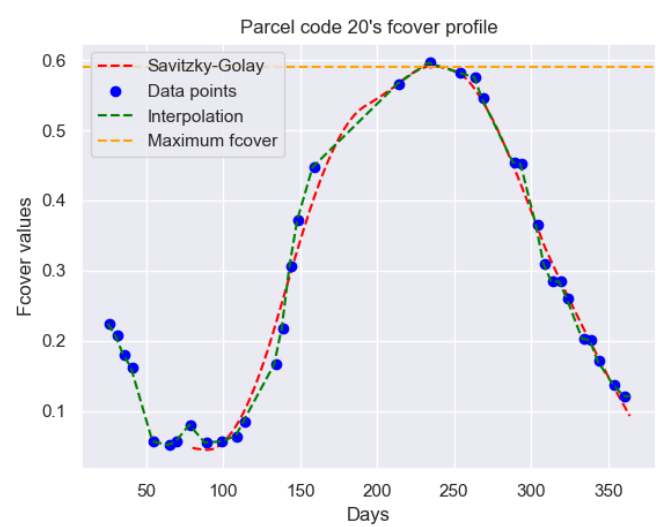
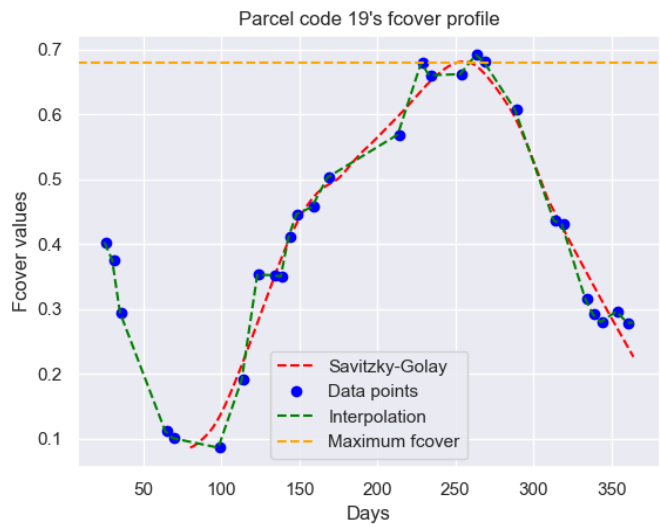
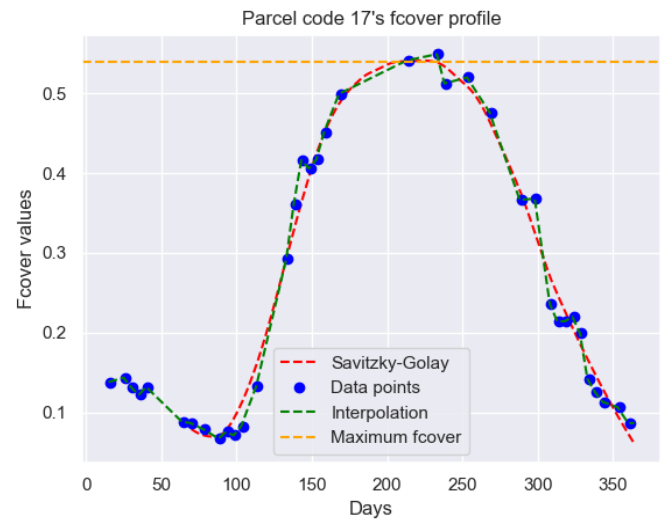
CHAPTER 8. Appendix

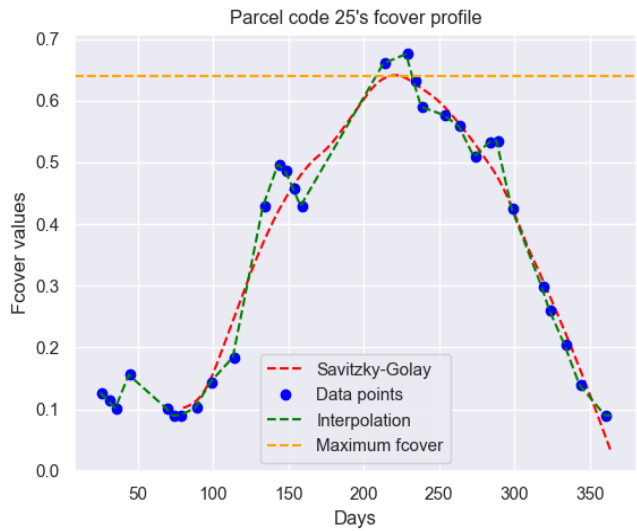
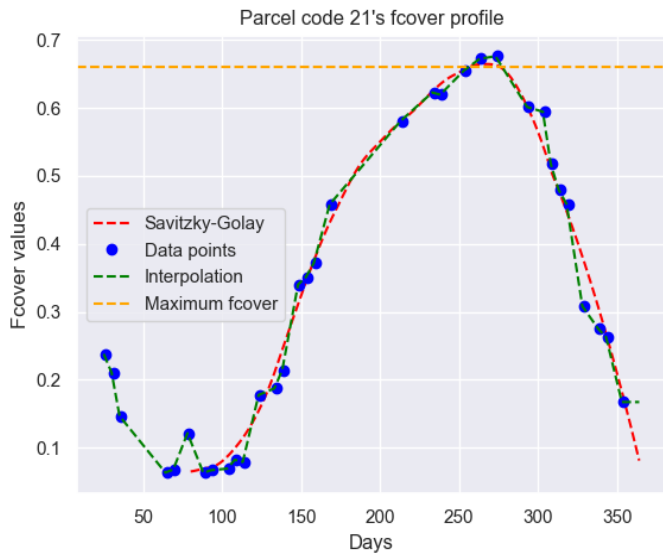
Appendix 1: Parcels' FCover profiles

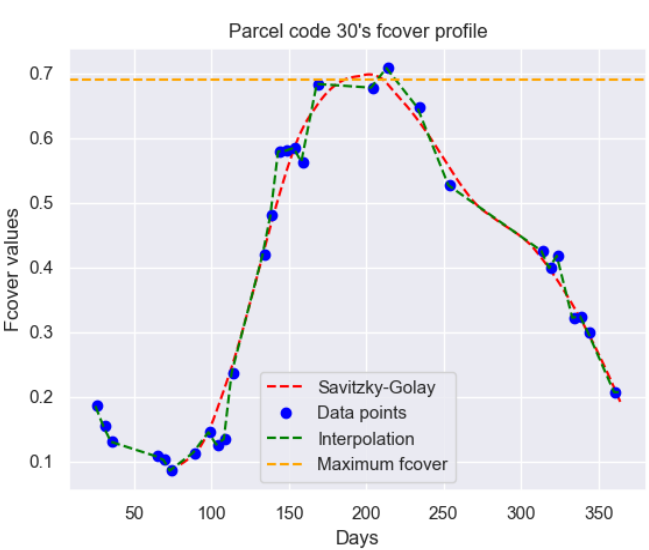
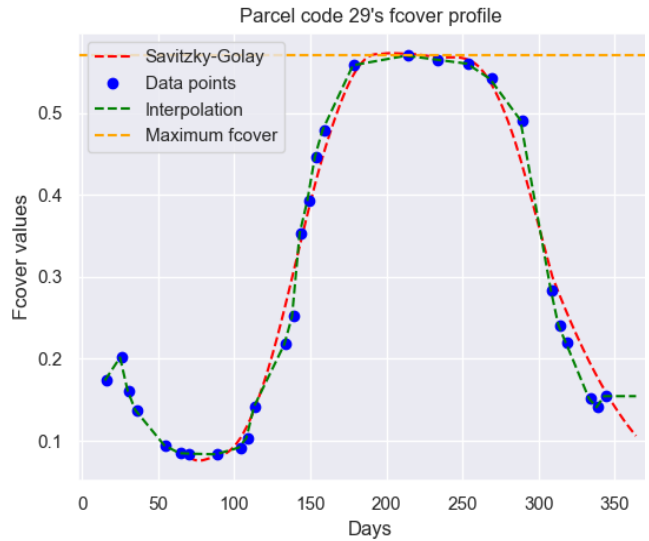


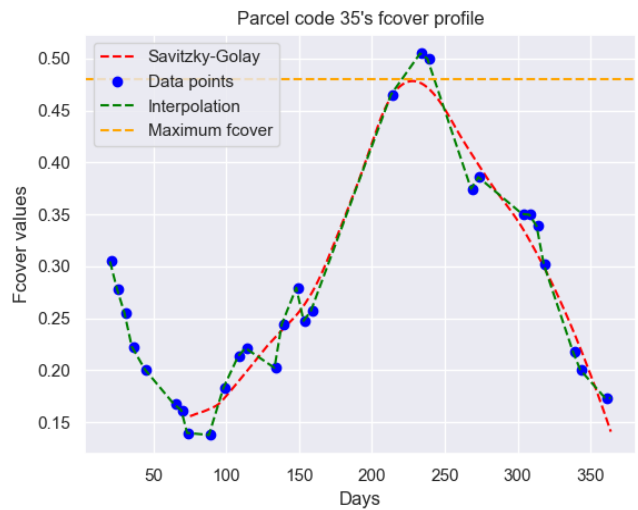
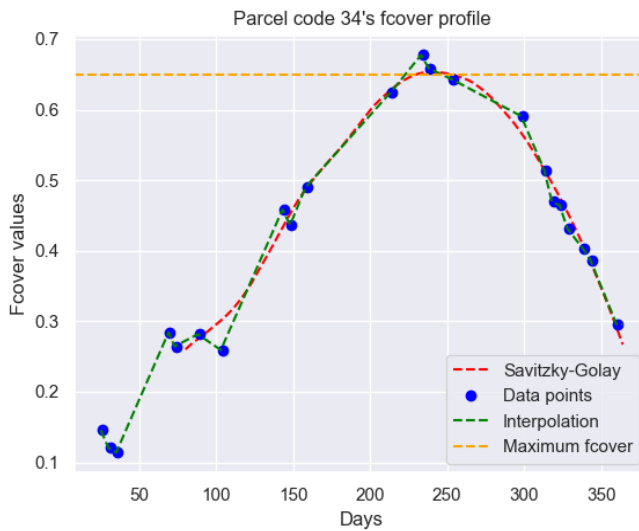
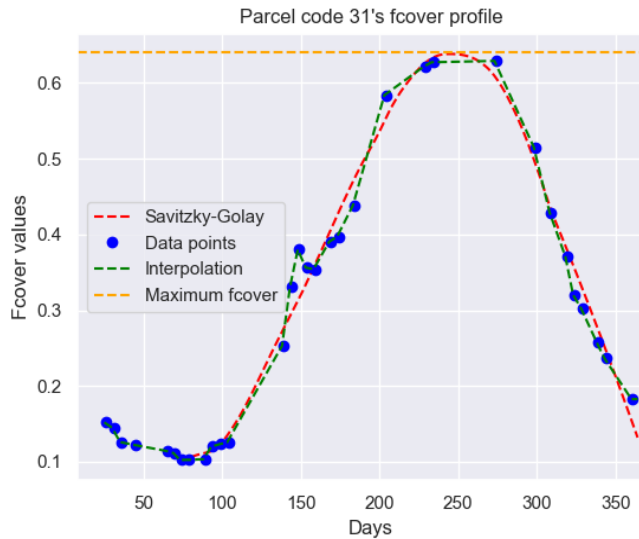


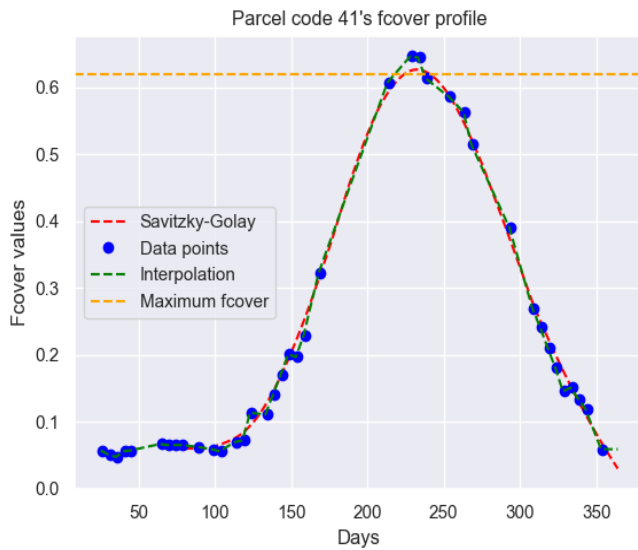
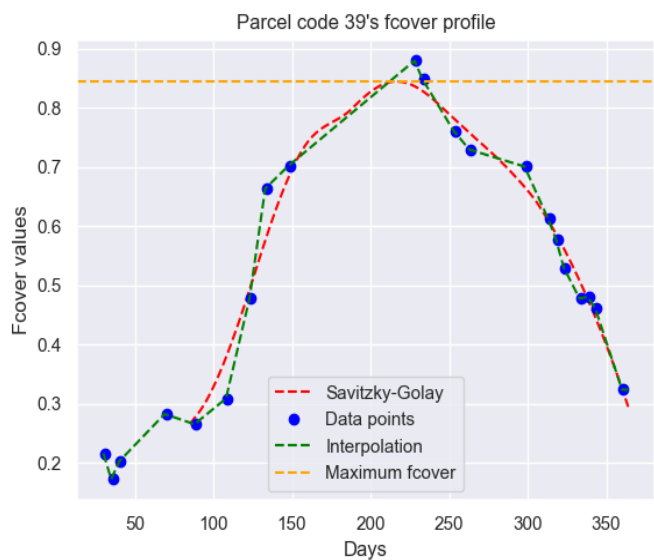
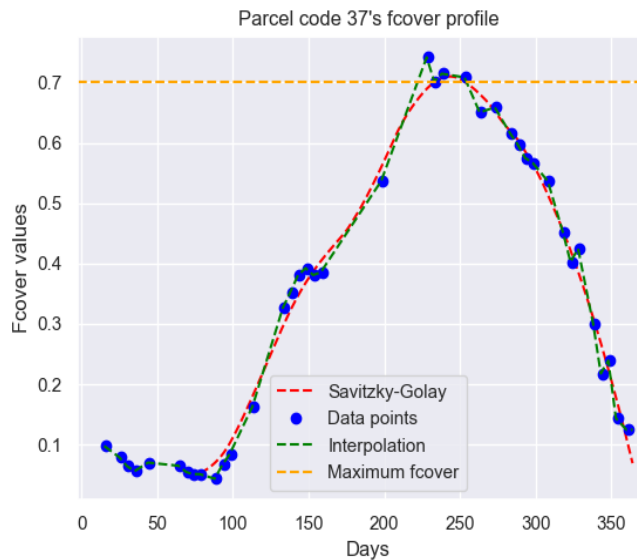


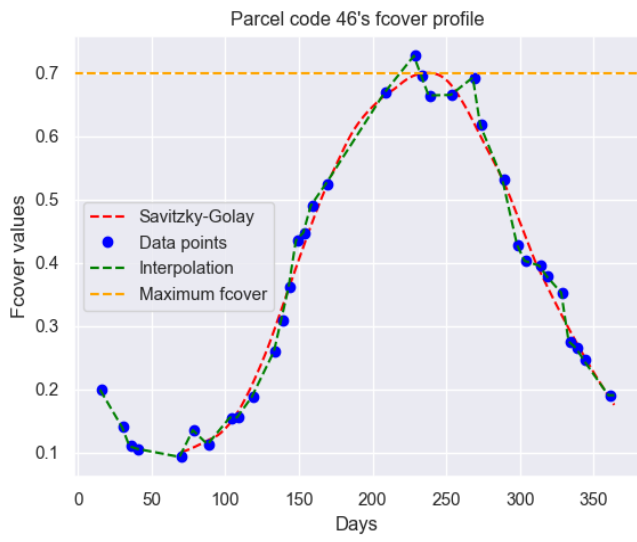
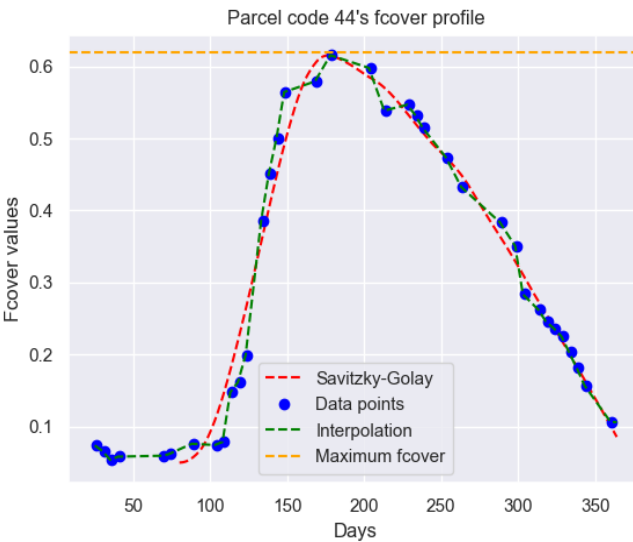
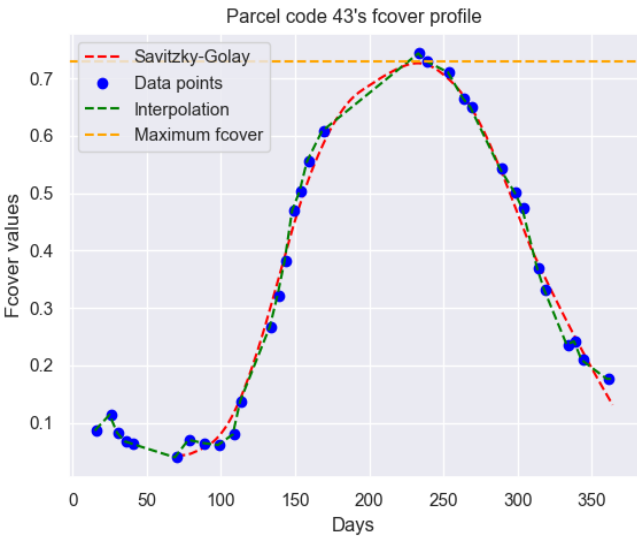
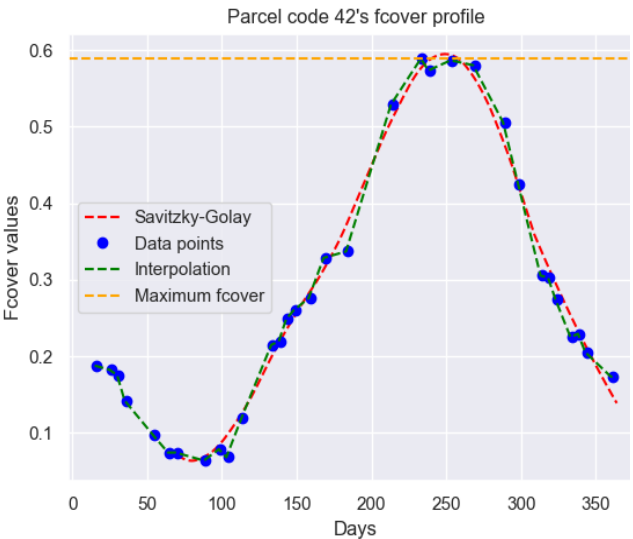


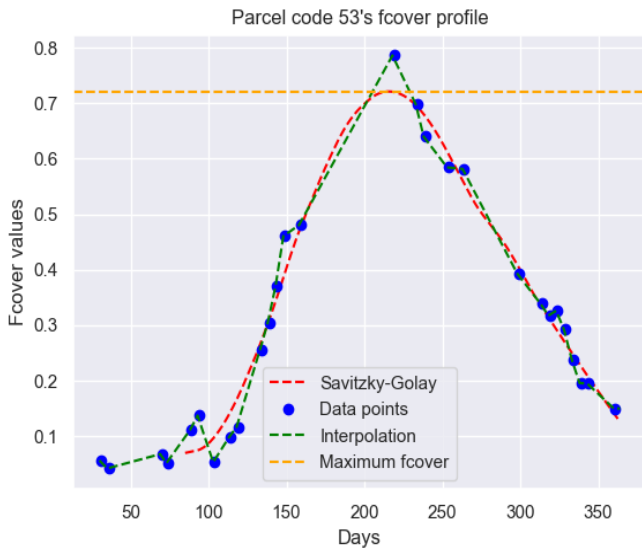
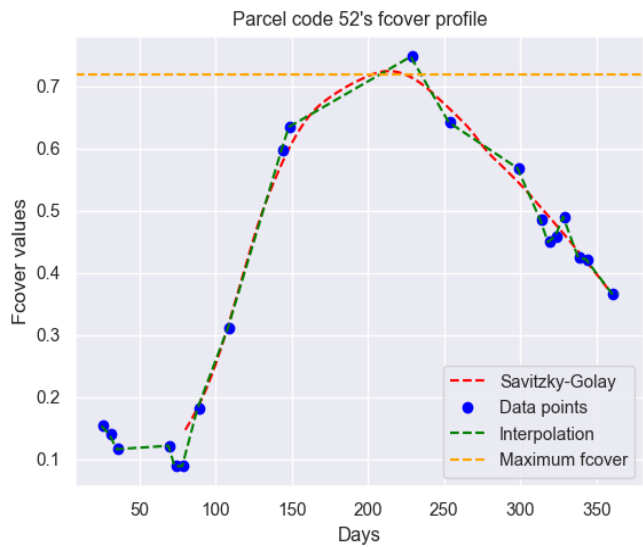
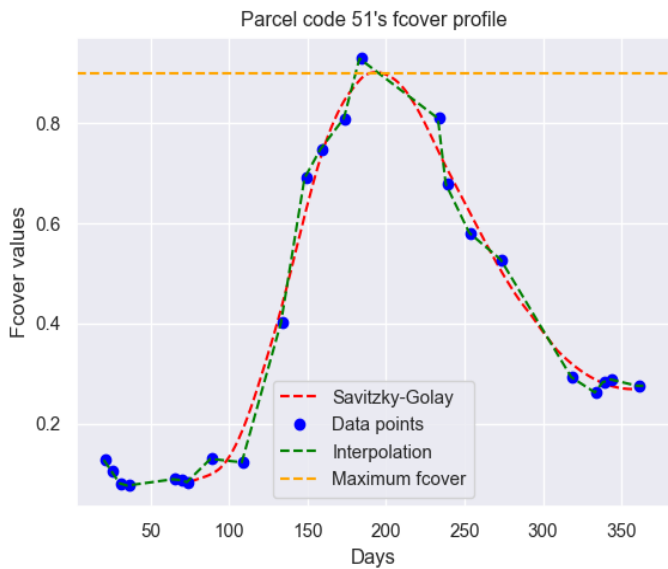


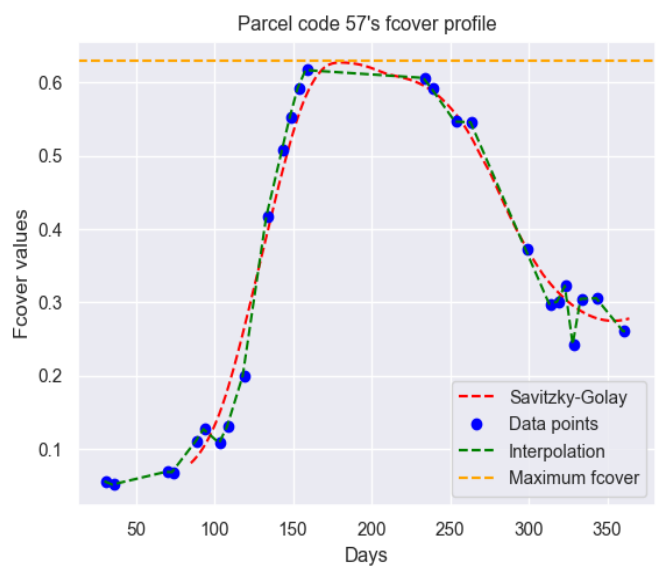
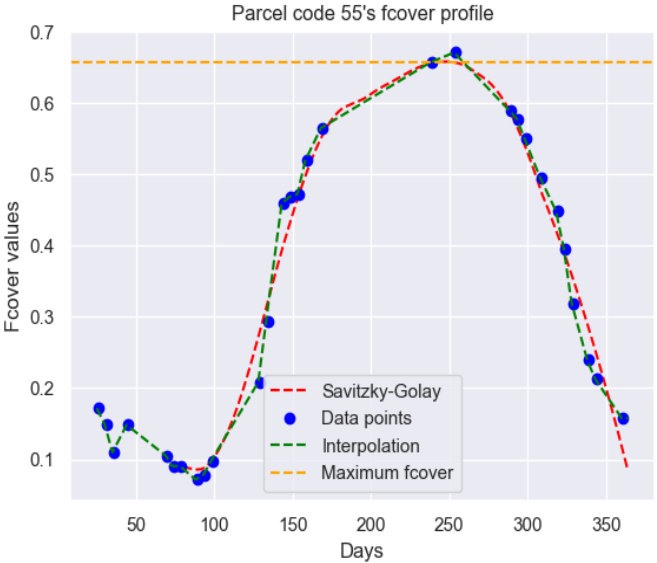
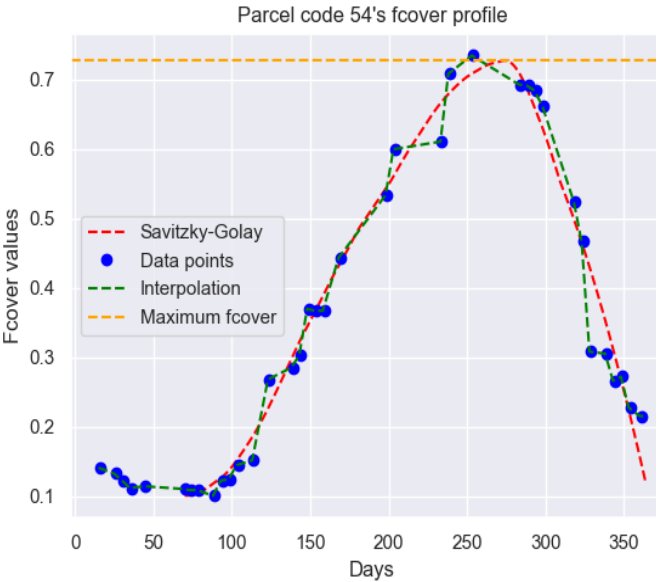


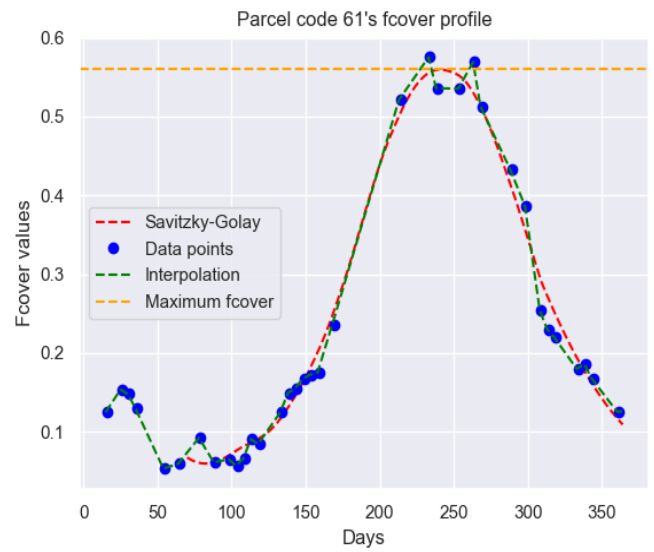
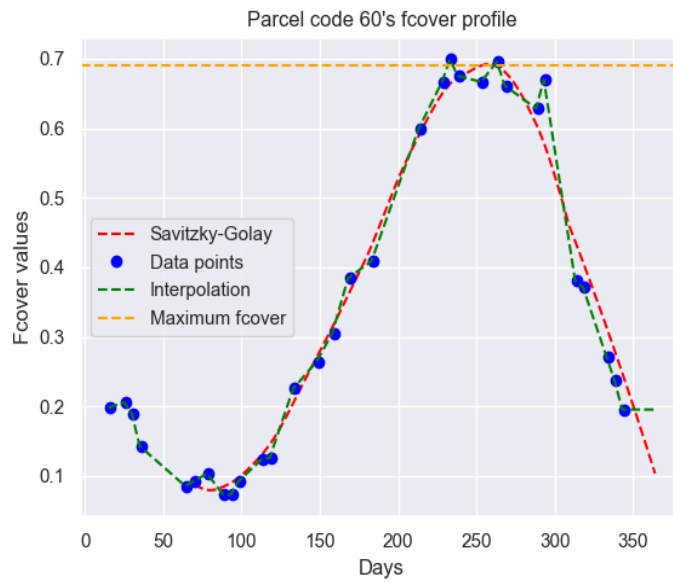
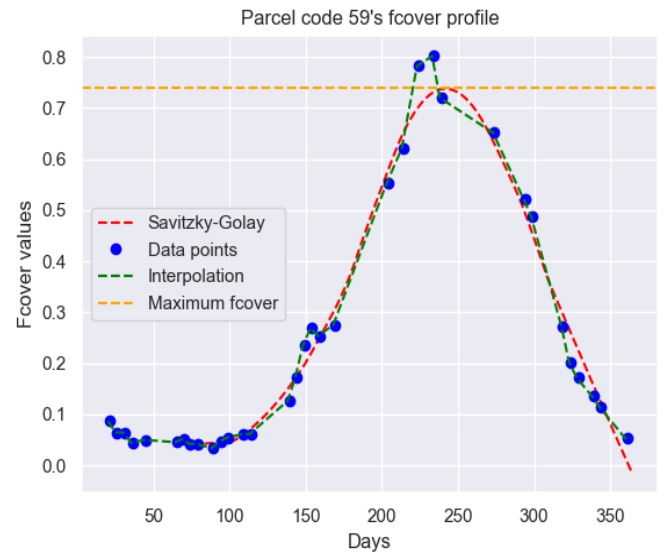
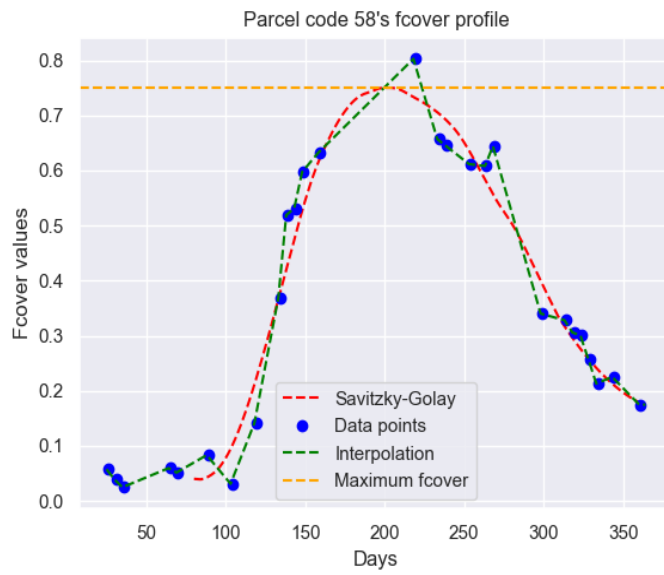




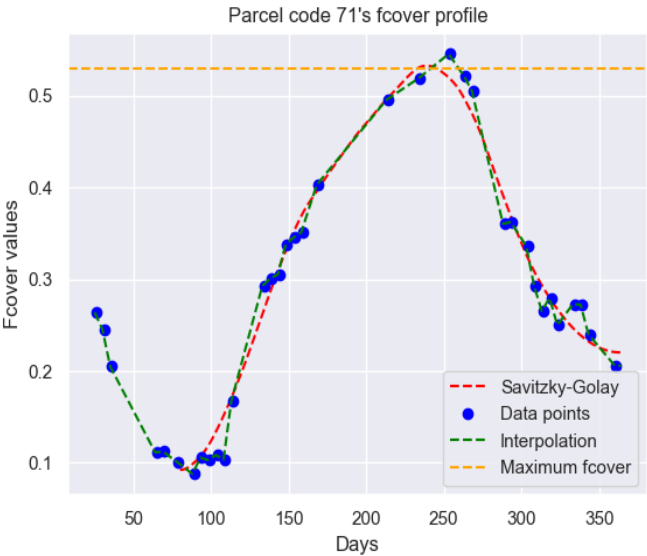
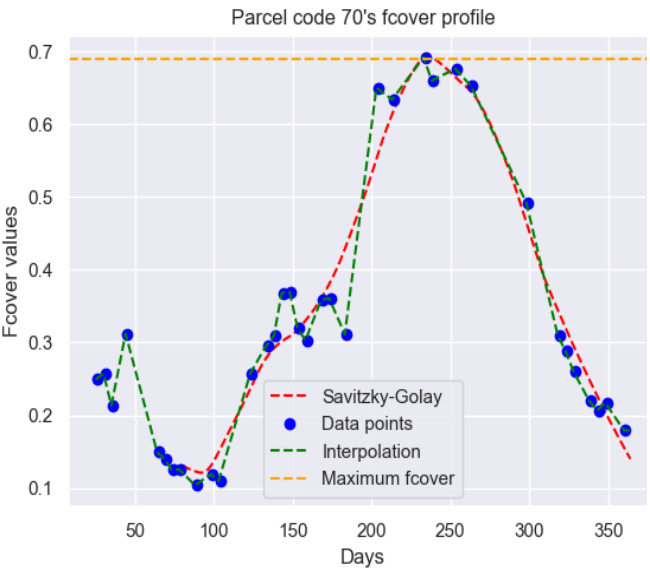
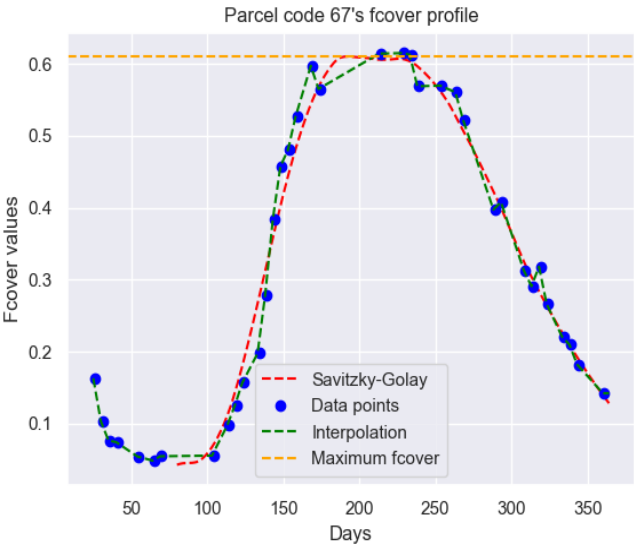
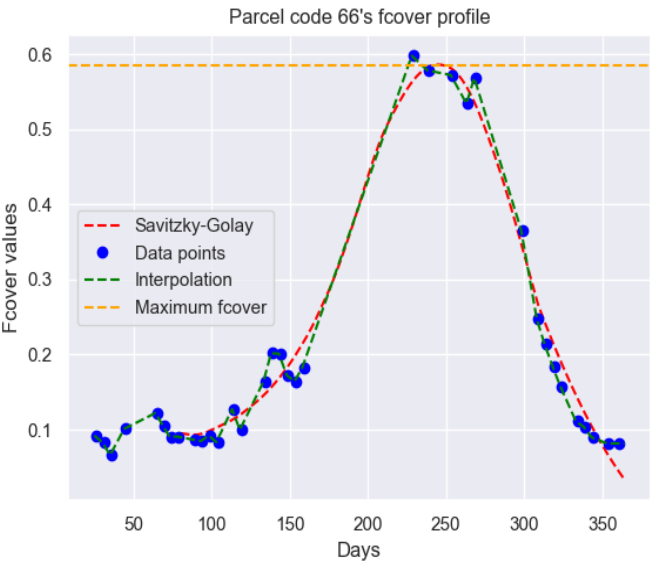


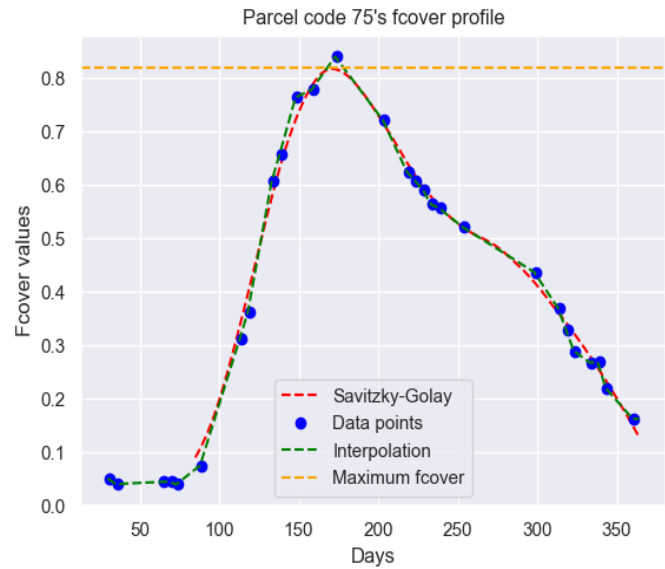
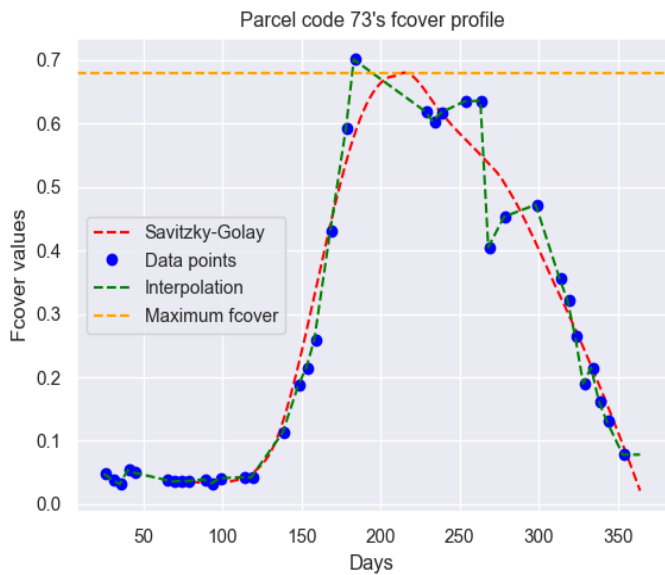
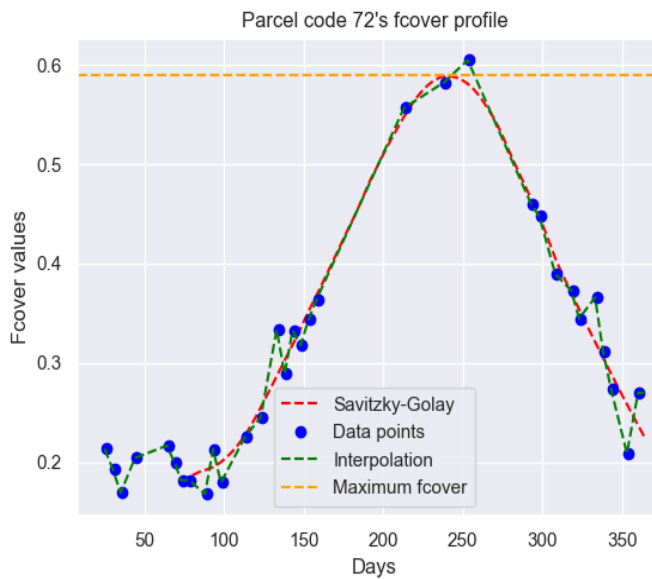


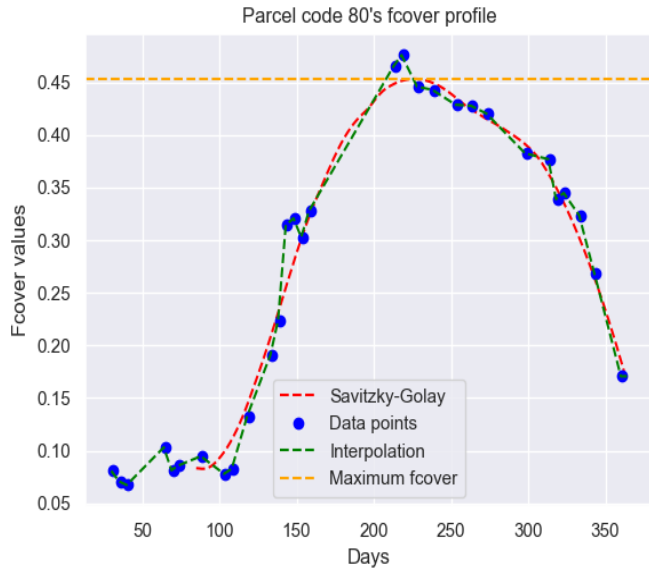
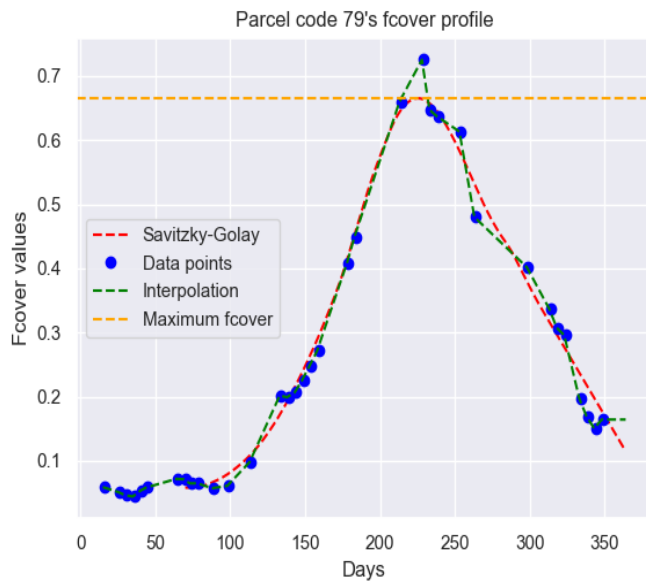
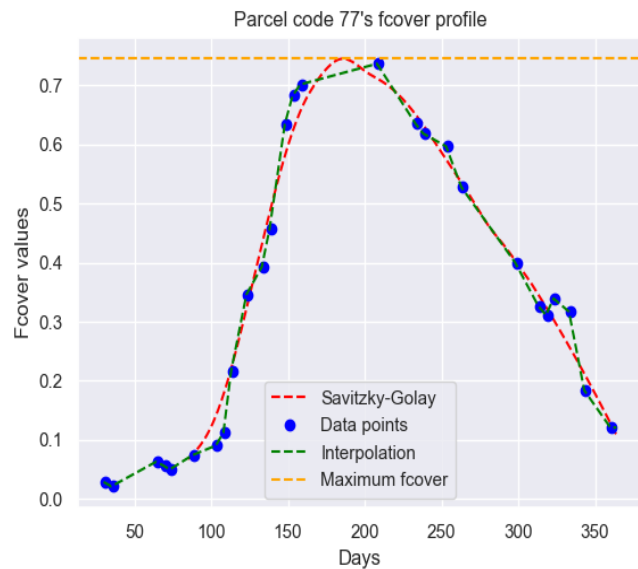
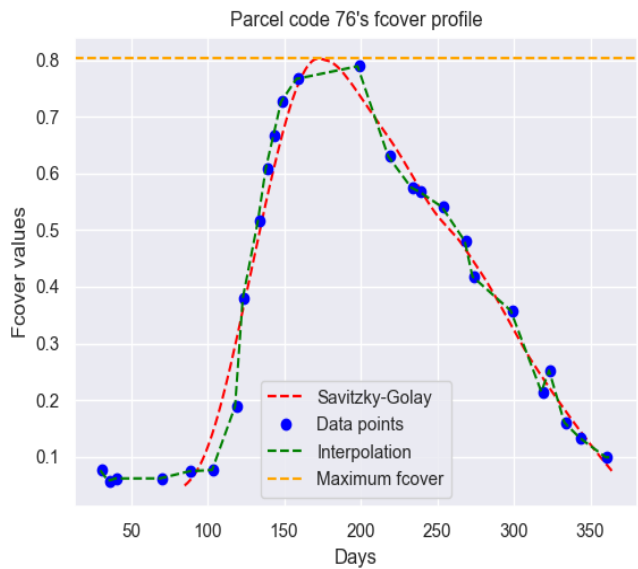


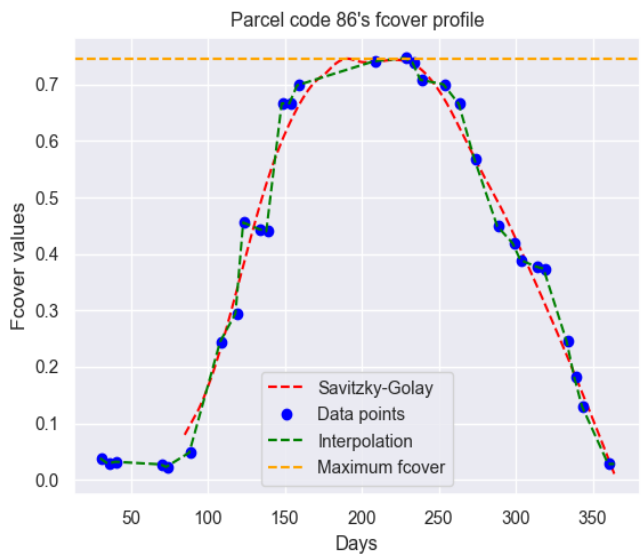
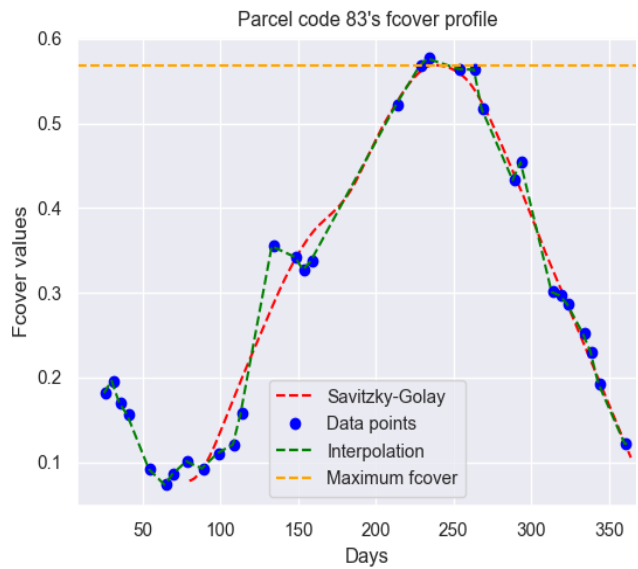
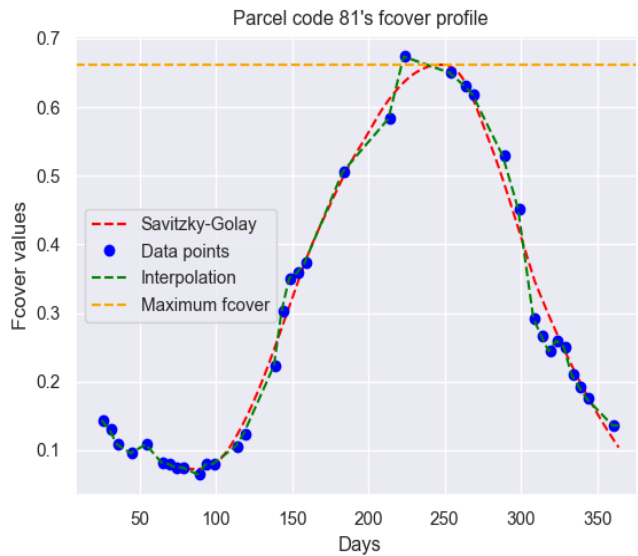


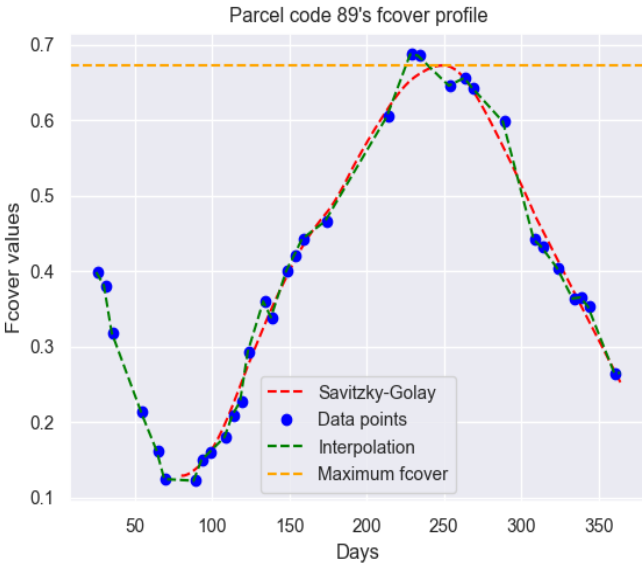
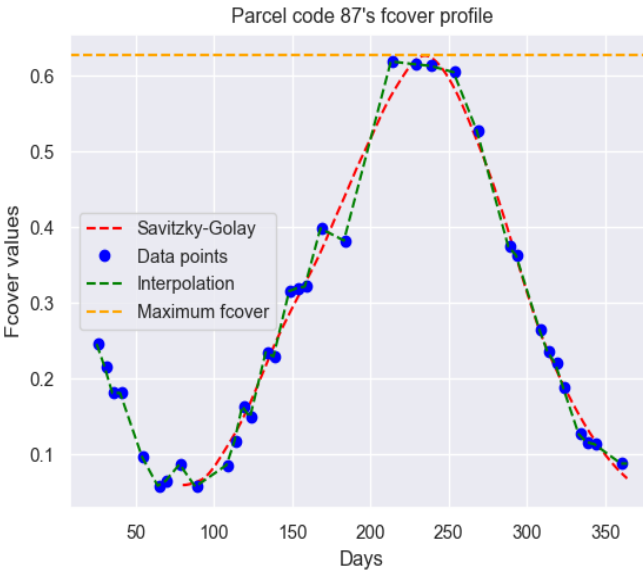


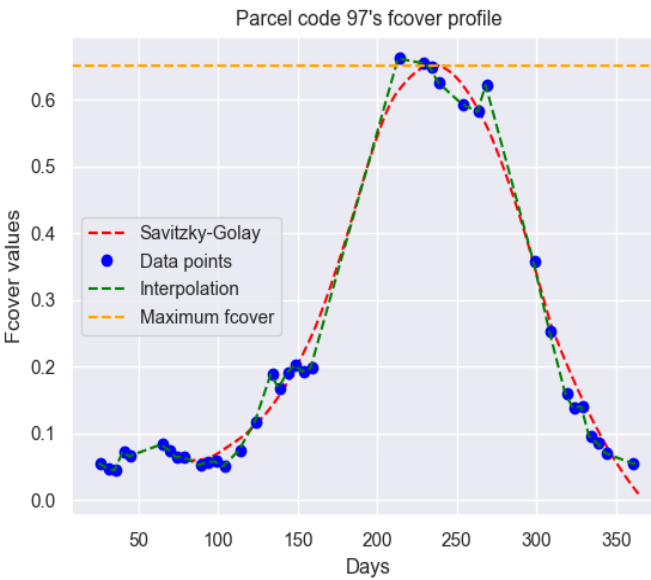
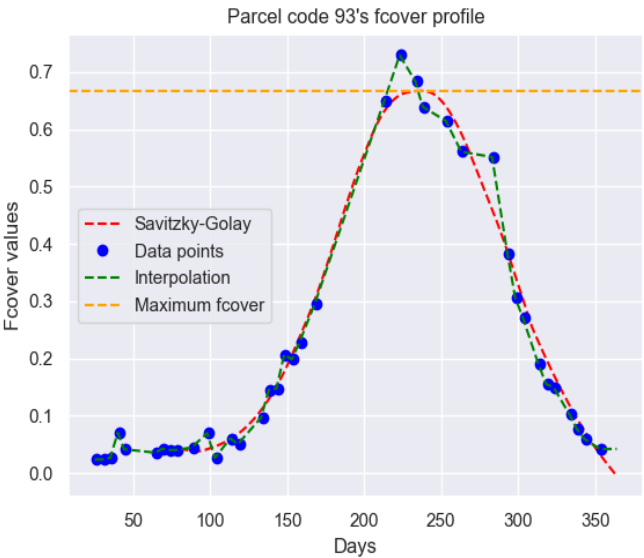


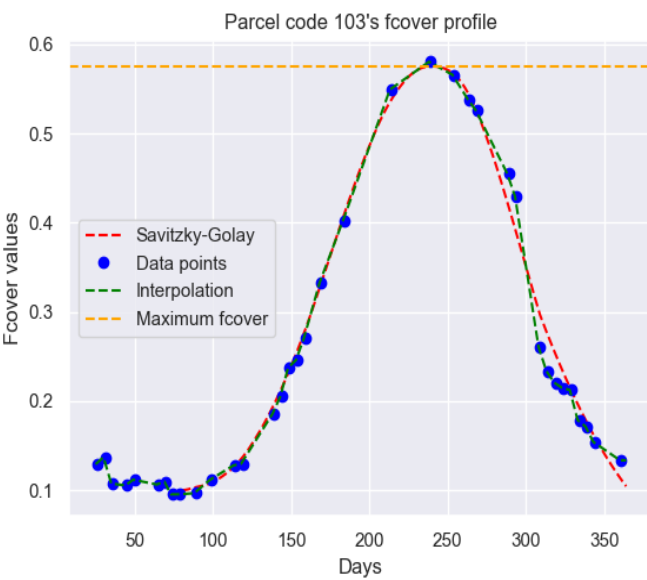
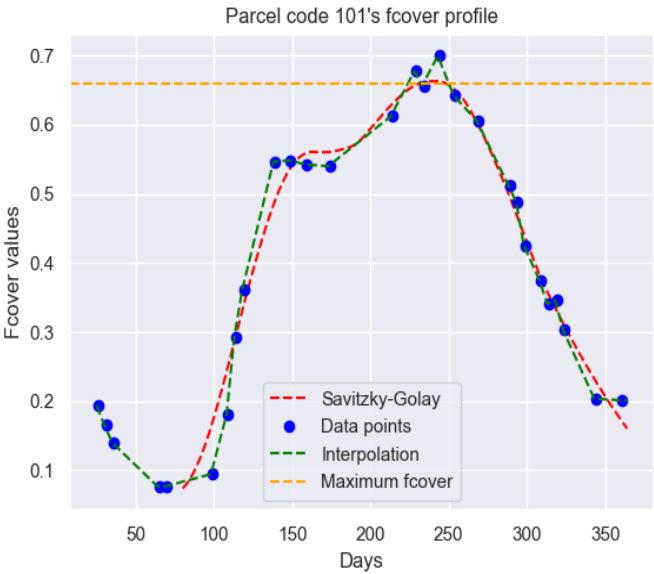
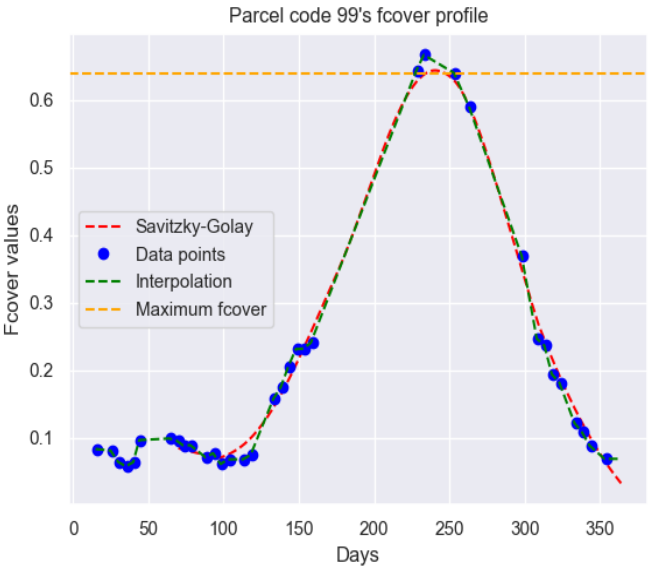


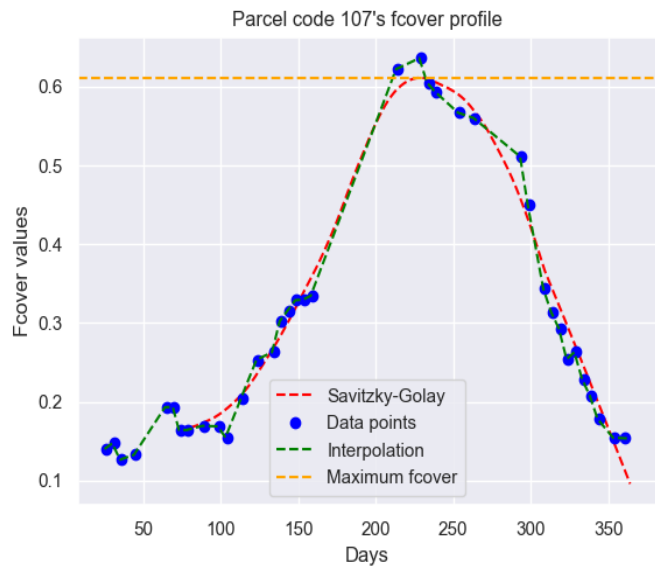
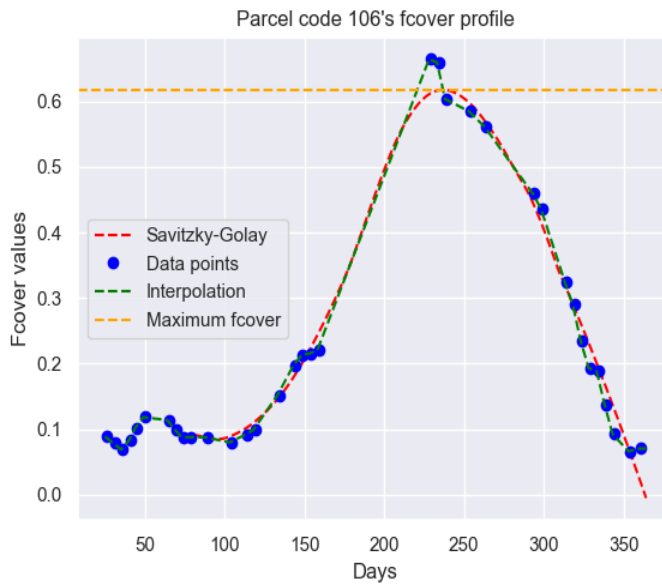
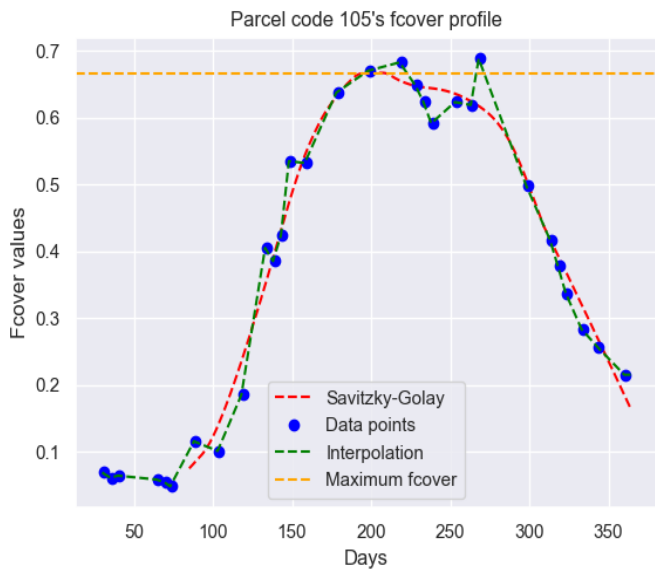
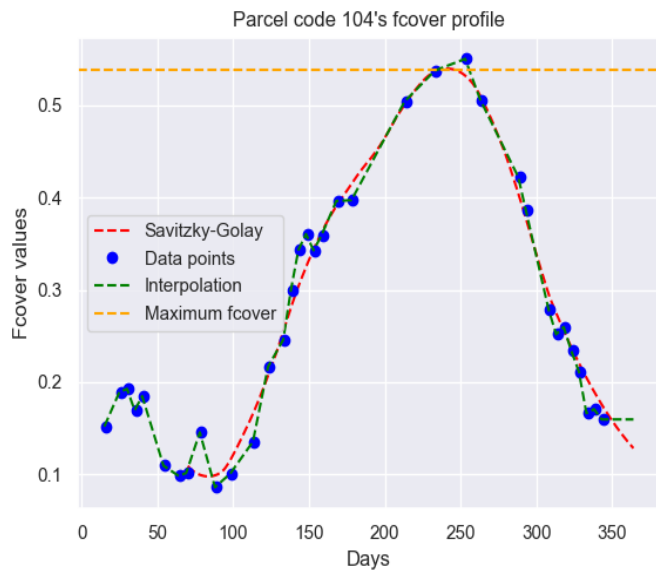


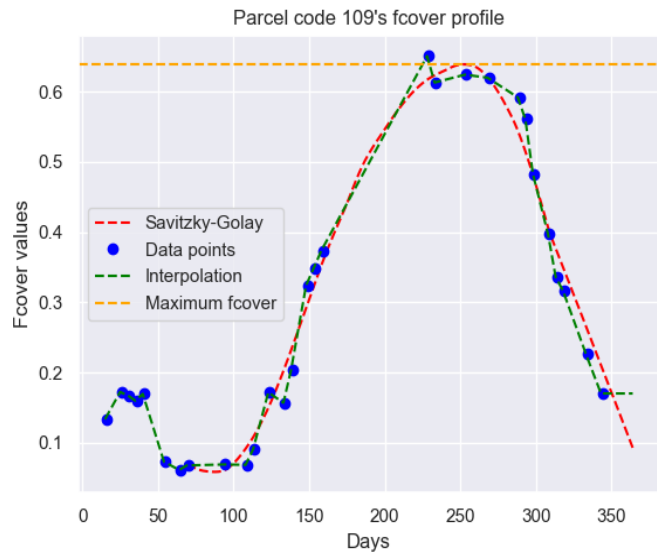


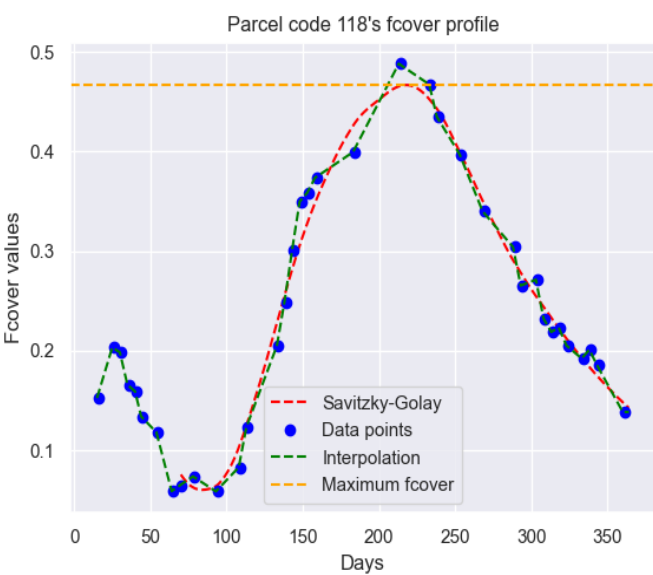
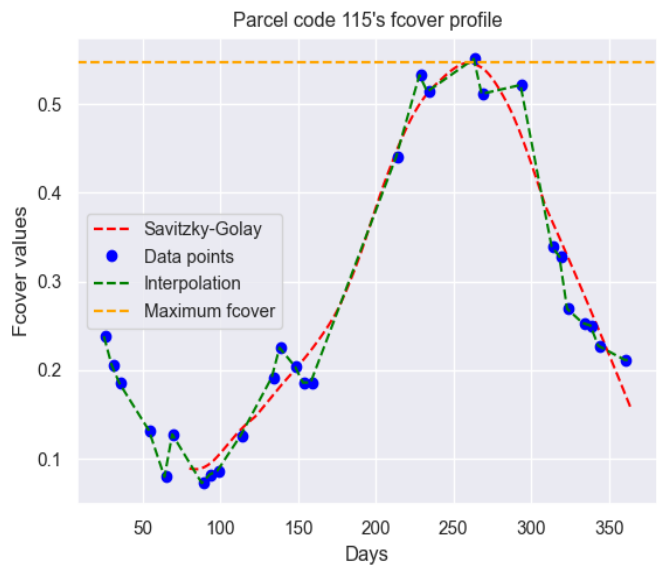
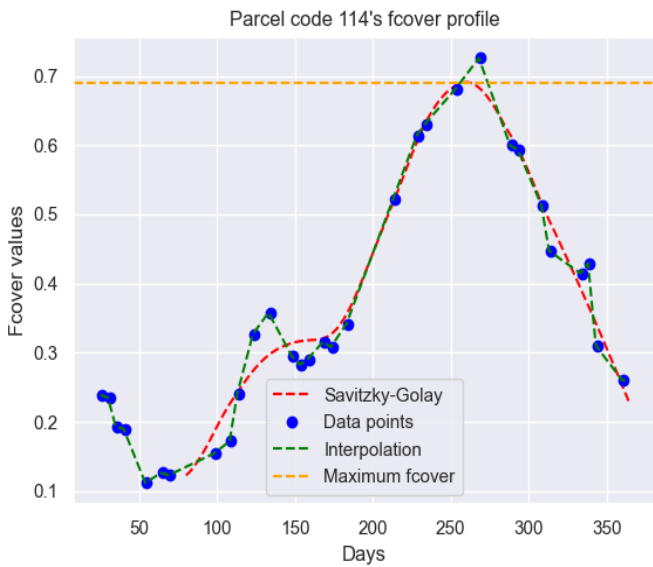
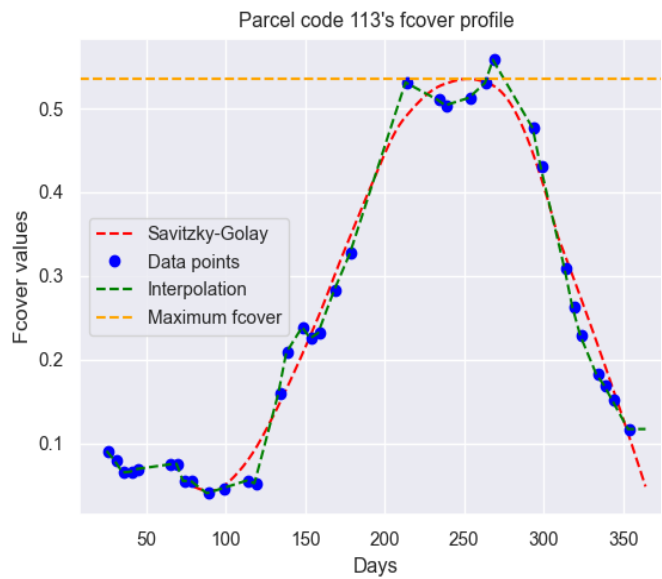


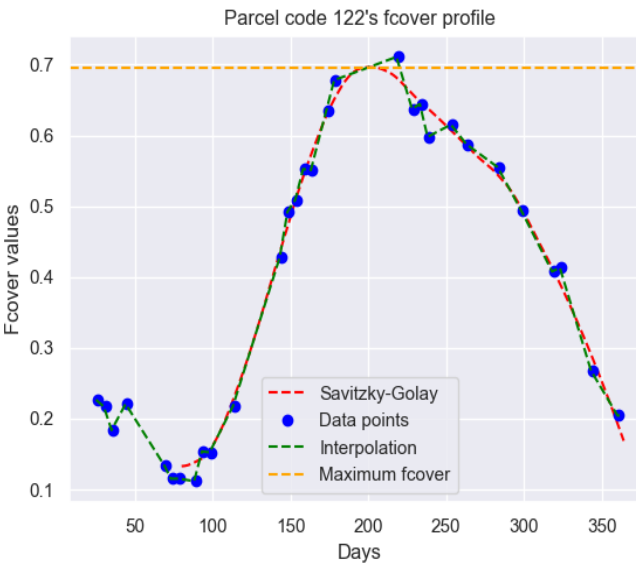
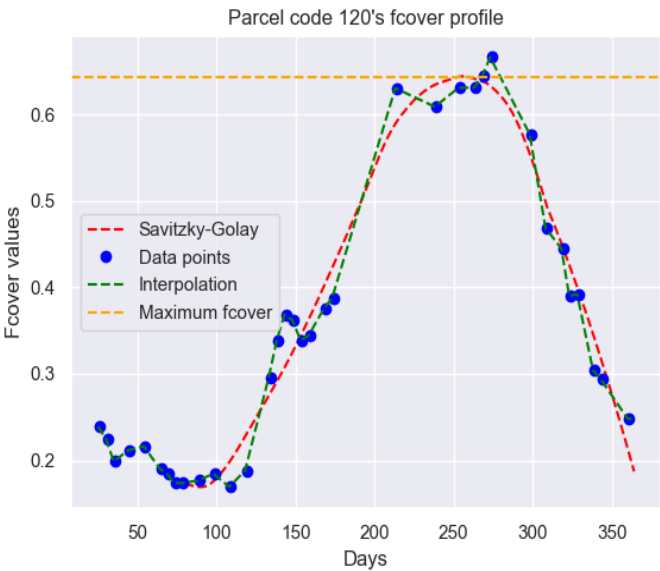
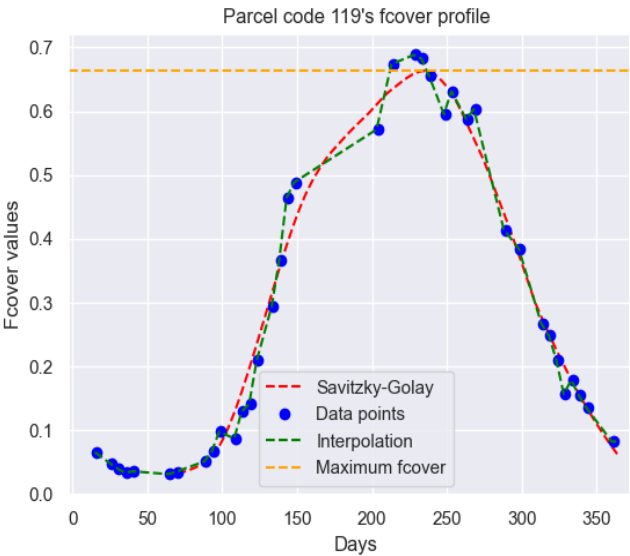


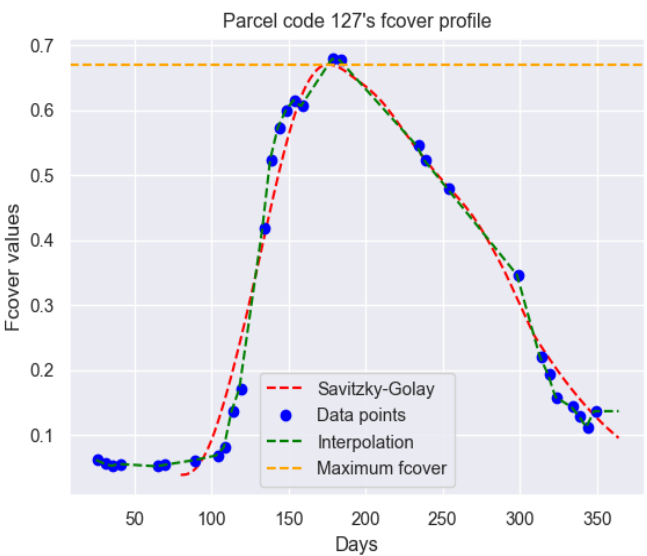
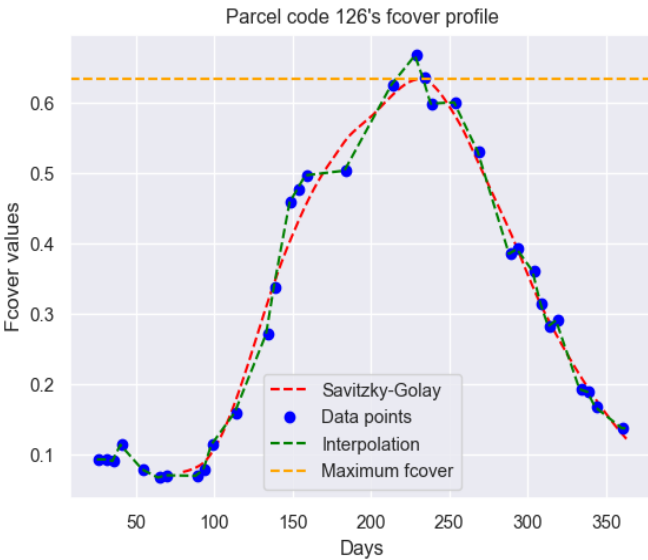
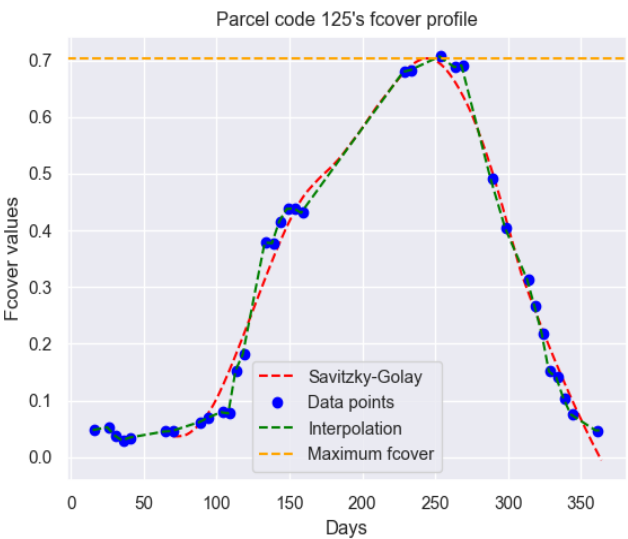
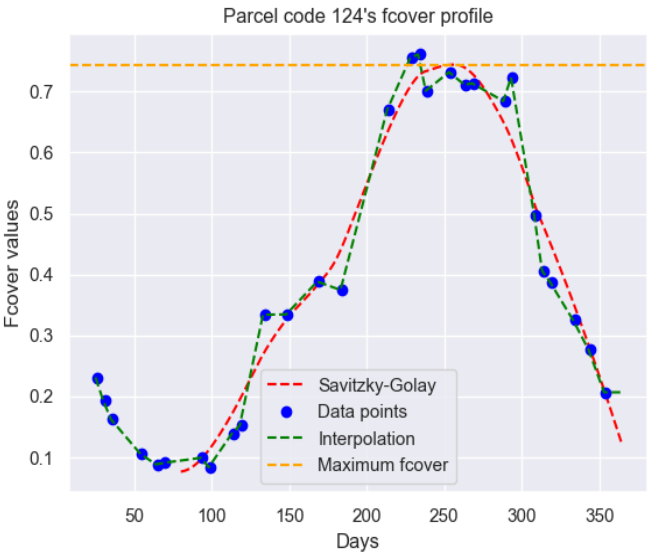


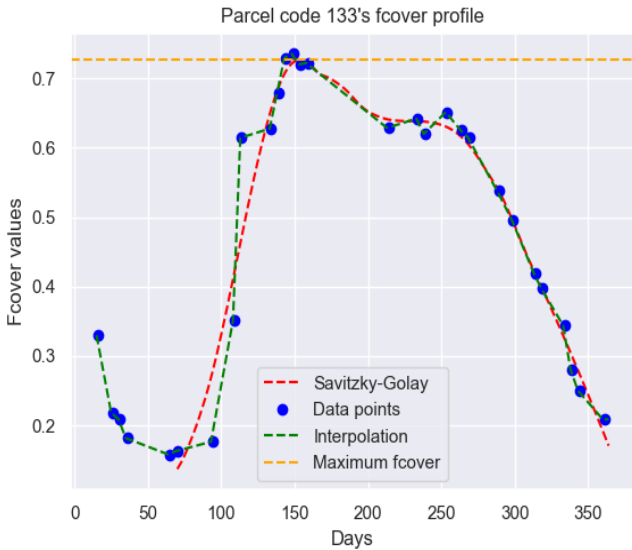
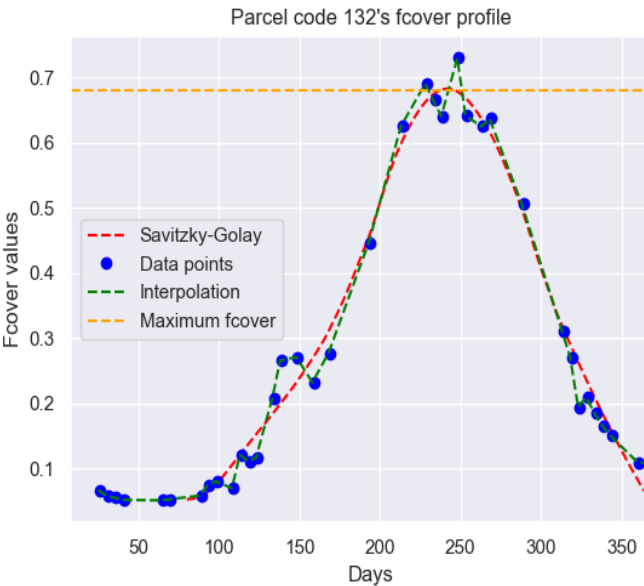
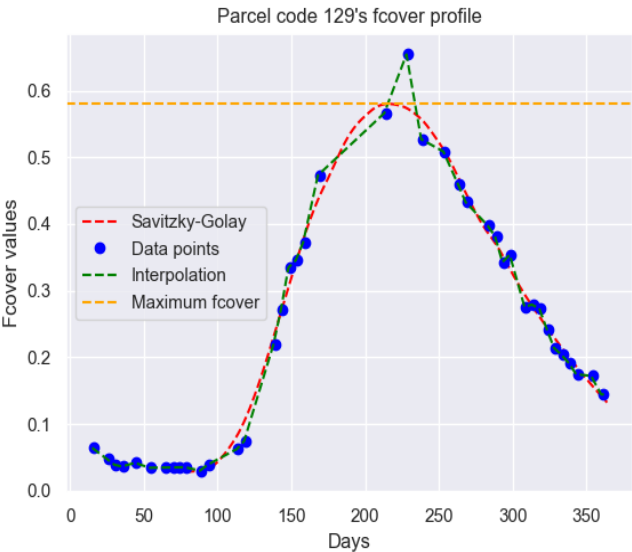
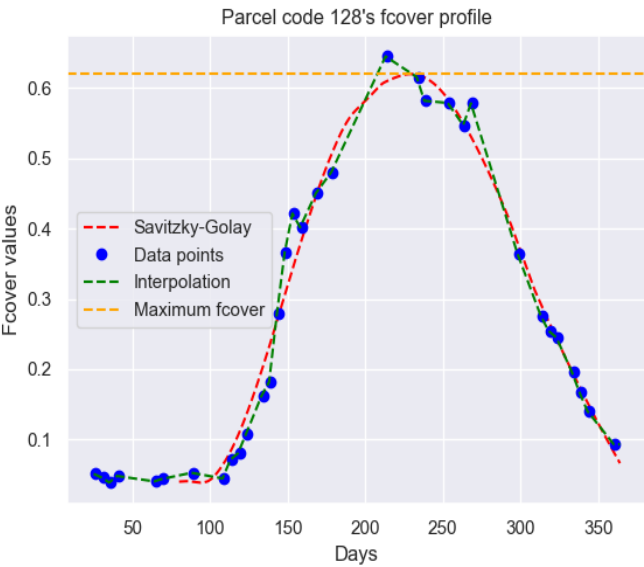


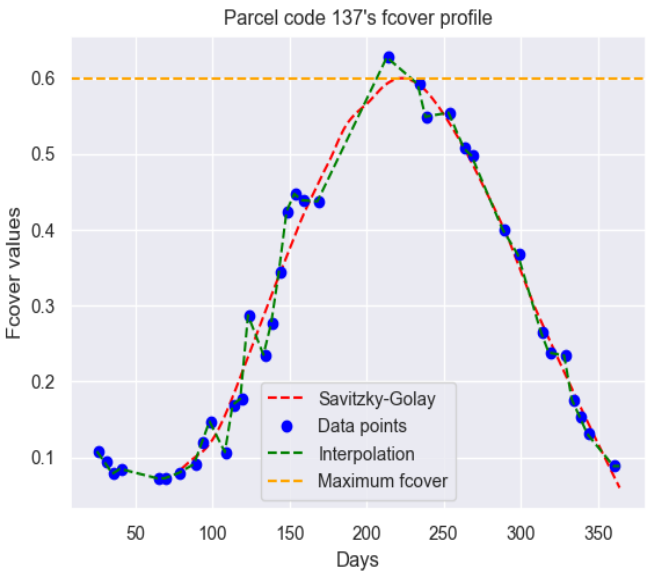
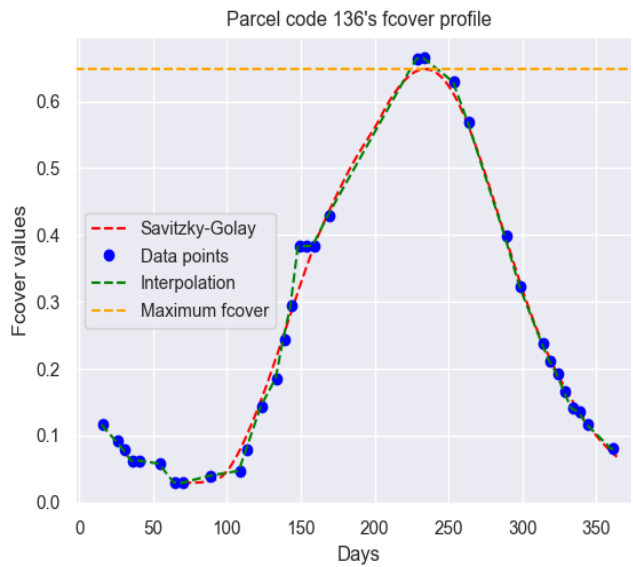
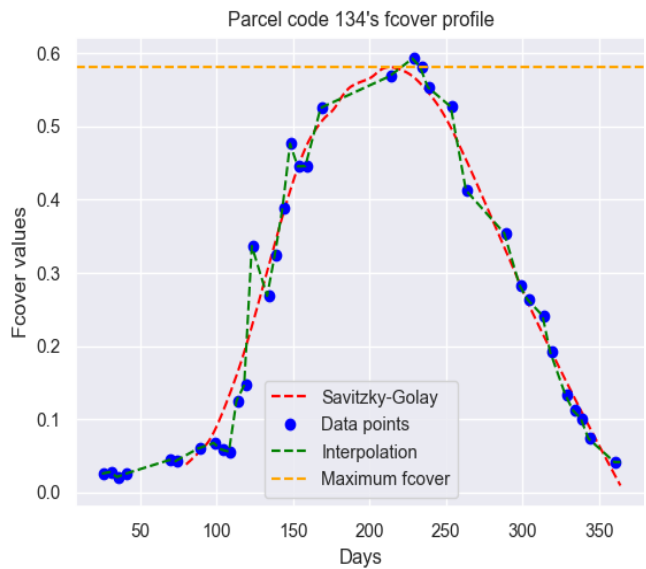


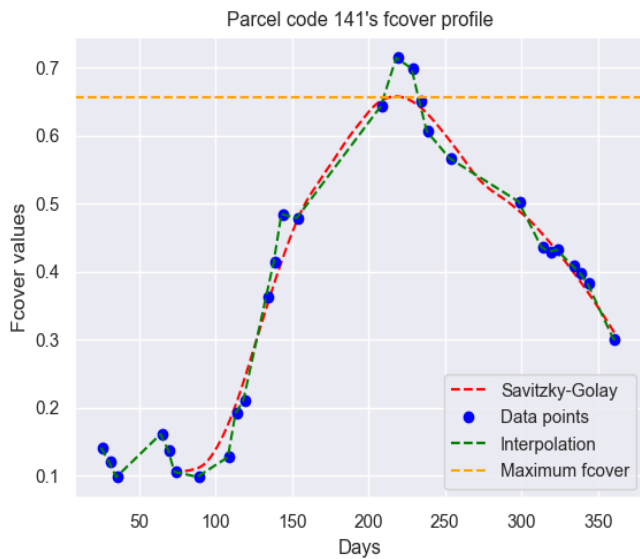
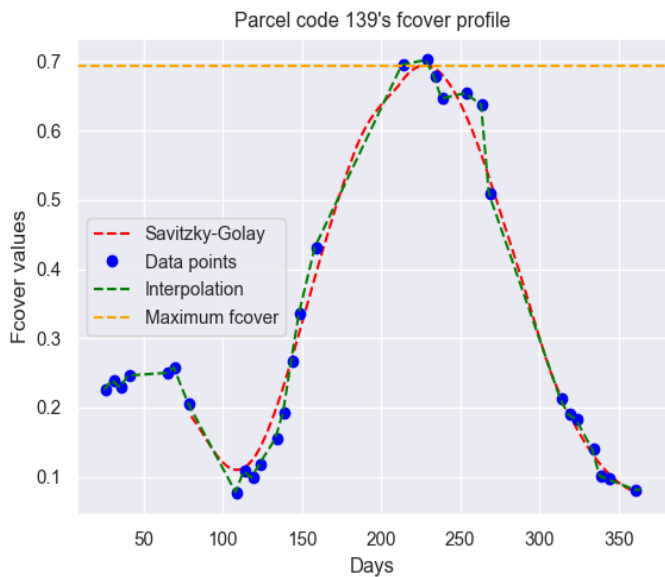
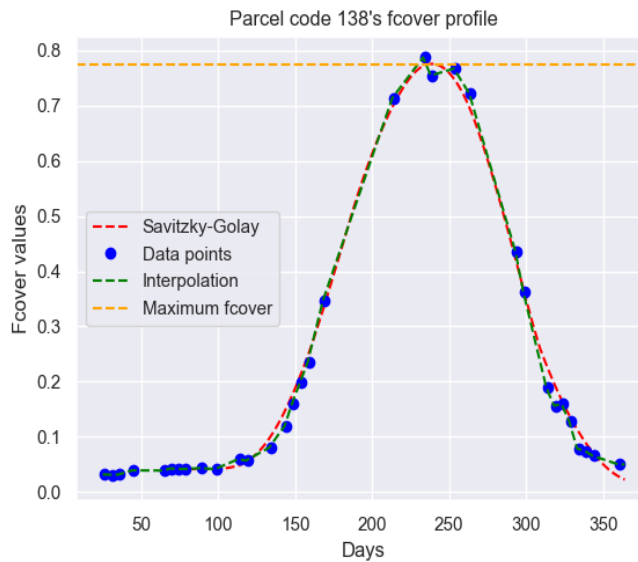












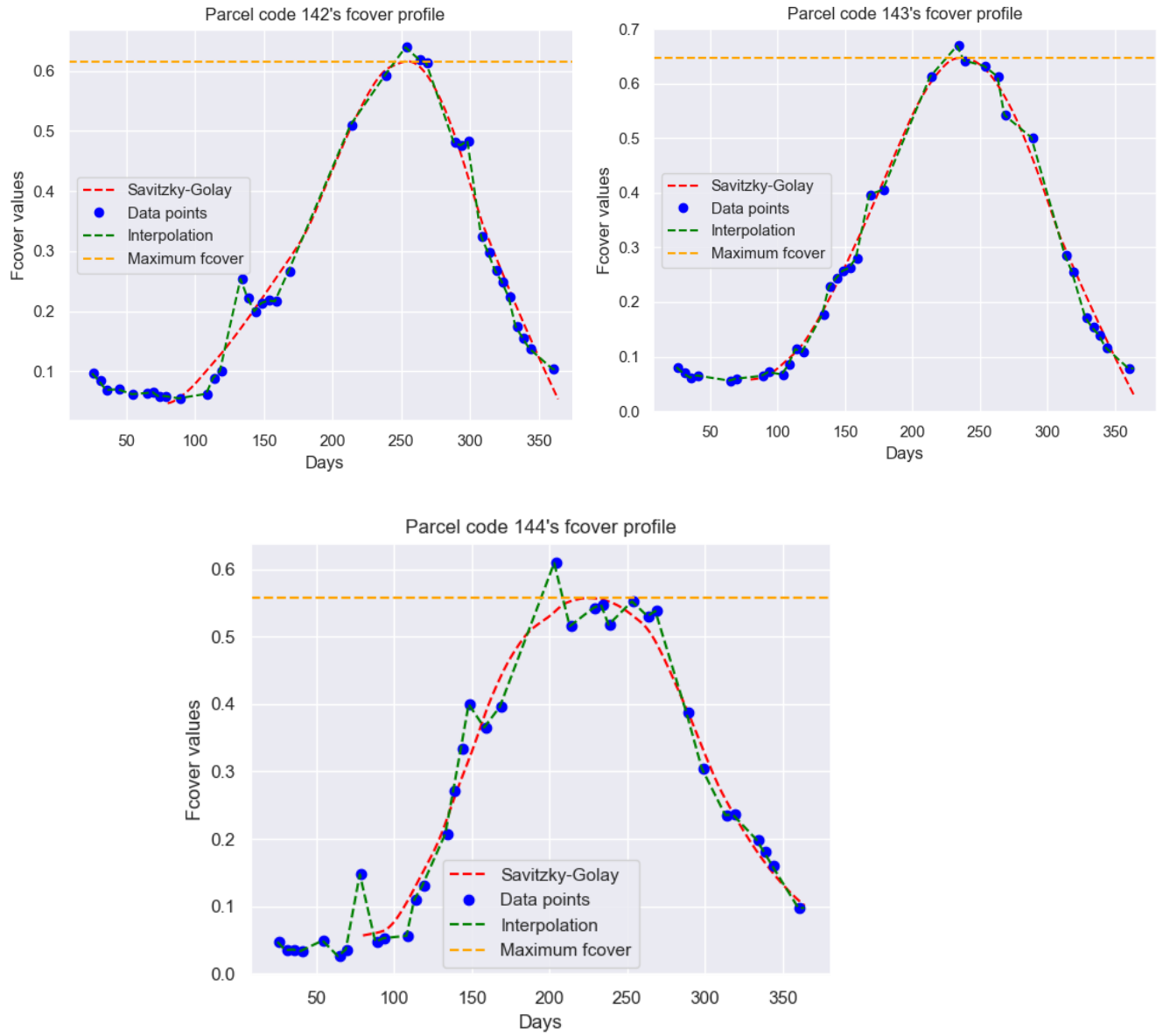


Figure 42: FCover profiles for all the parcels

Appendix 2: Elevation Map of Keiyo Marakwet and Uasin Gishu

The Elevation dataset was obtained from Copernicus GLO-30, which offers global coverage at a resolution of 30 meters provided in DEM format. The dataset was used to generate the map shown in Figure 43, showing the difference in elevation within the area of interest. Uasin Gishu and Keiyo Marakwet possess contrasting landscapes (see Figure 44 under Appendix 3), with most regions in Uasin Gishu lying on a plain. Keiyo Marakwet is characterized by lowlands to the East, highlands to the Northwest and mid-altitude areas on the Western part.

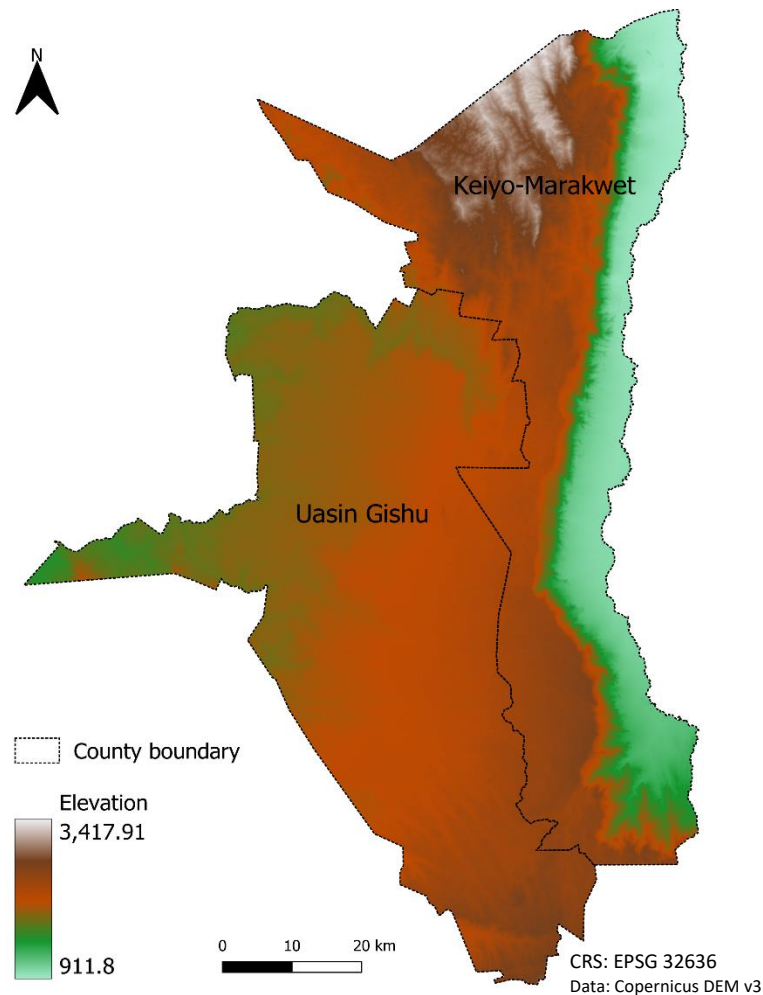


Figure 43: Elevation map of Uasin Gishu and Keiyo Marakwet

Appendix 3: Uasin Gishu and Keiyo Marakwet landforms

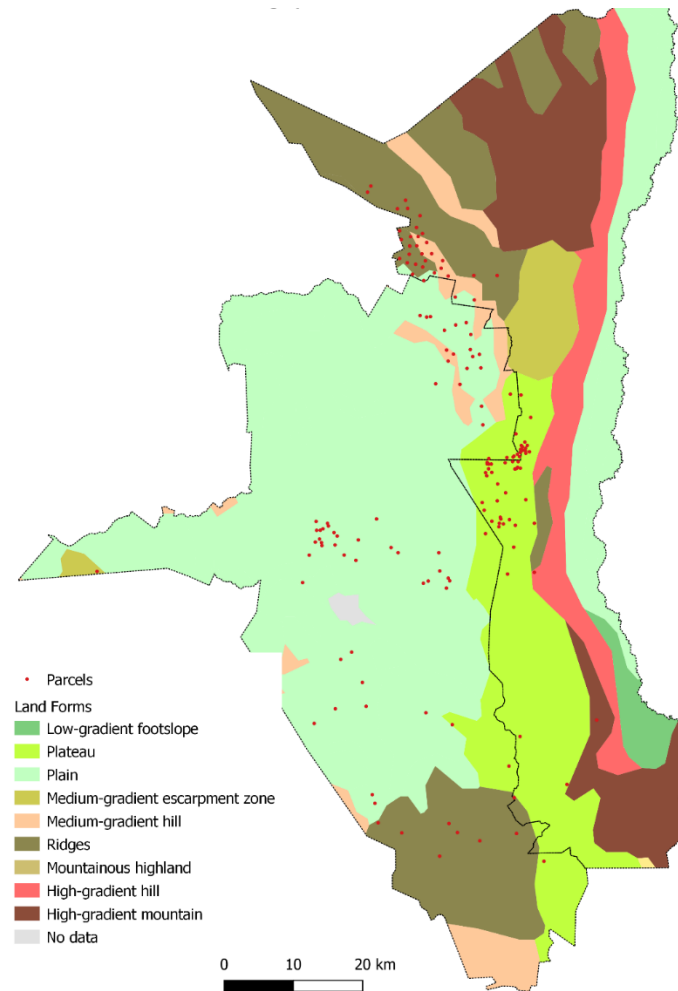


Figure 44: Landforms in Uasin Gishu and Keiyo Marakwet showing the heterogeneous landscapes

Appendix 5: Model performance before assimilation according to planting dates

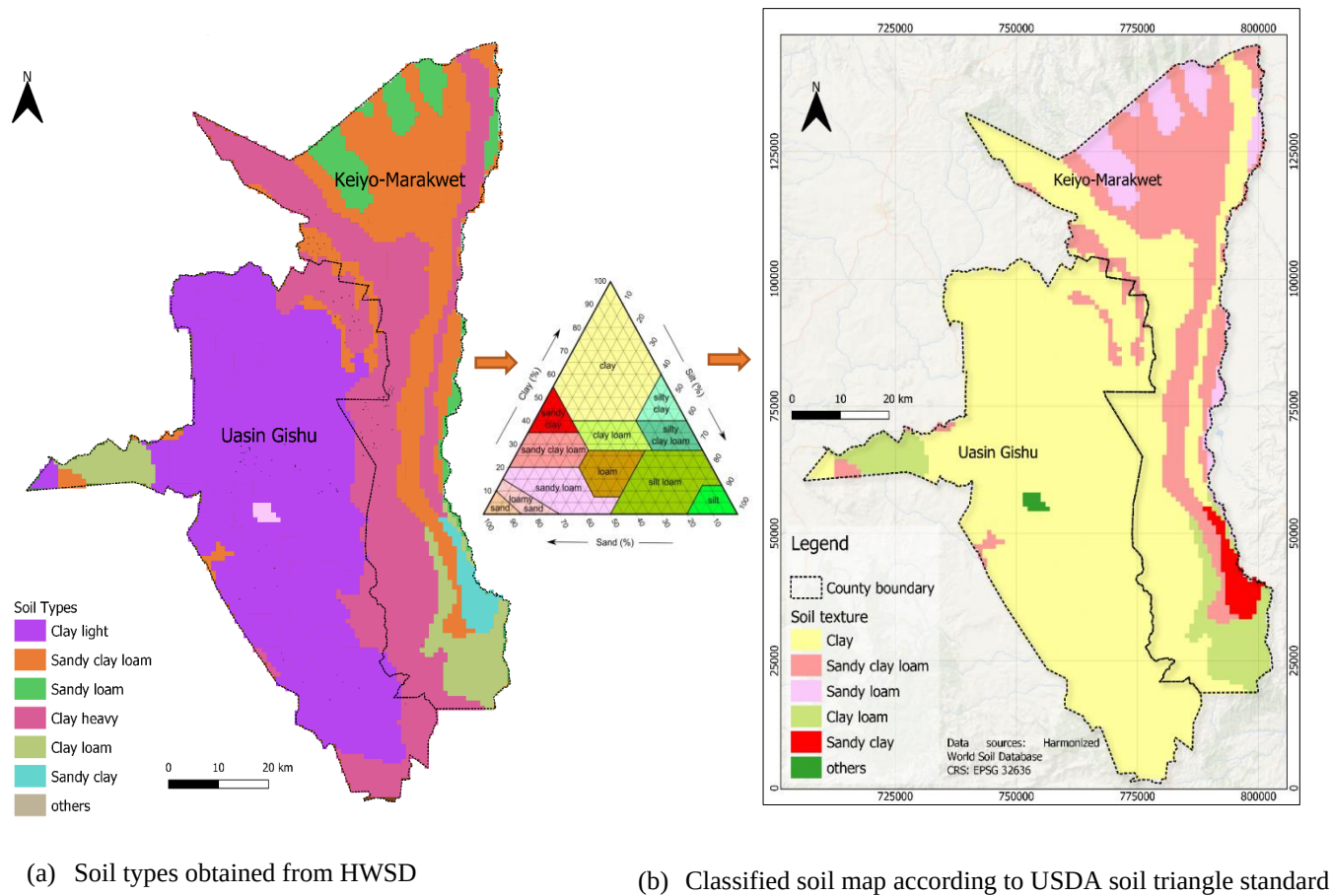


Figure 45: Soil texture classification

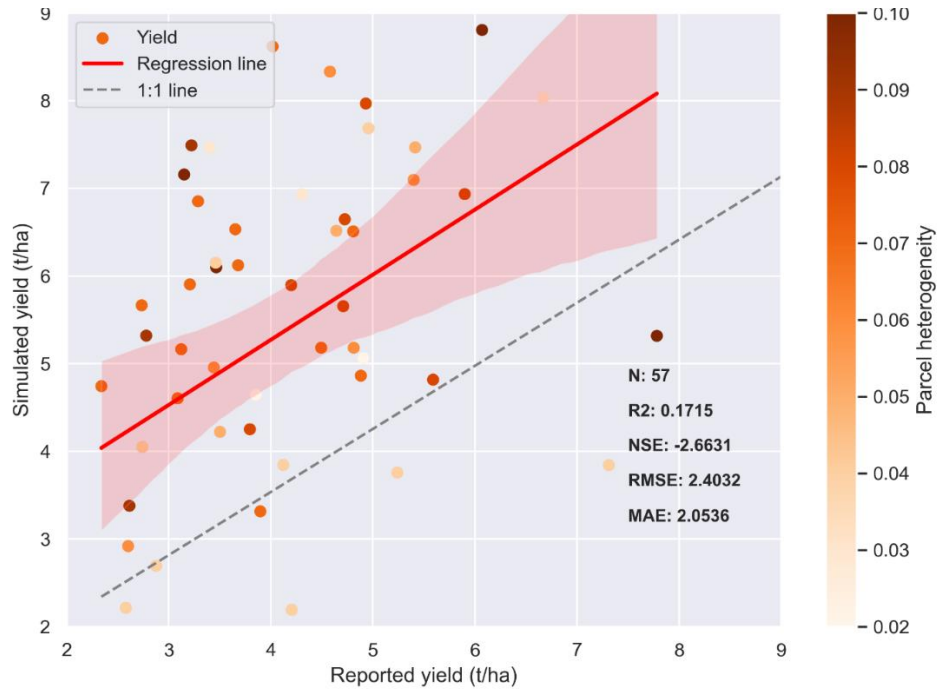


Figure 46: Model yield estimation performance before assimilation for March 27-14 April planting dates

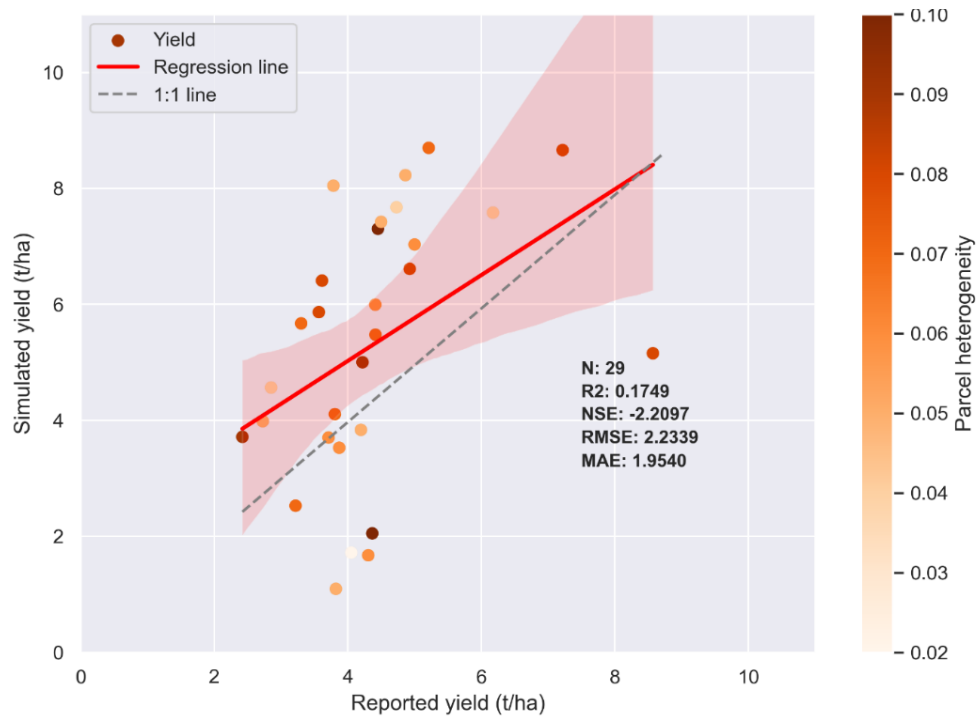


Figure 47: Model yield estimation performance before assimilation for April 15-27 planting dates

Appendix 6: Model performance after assimilation according to planting dates

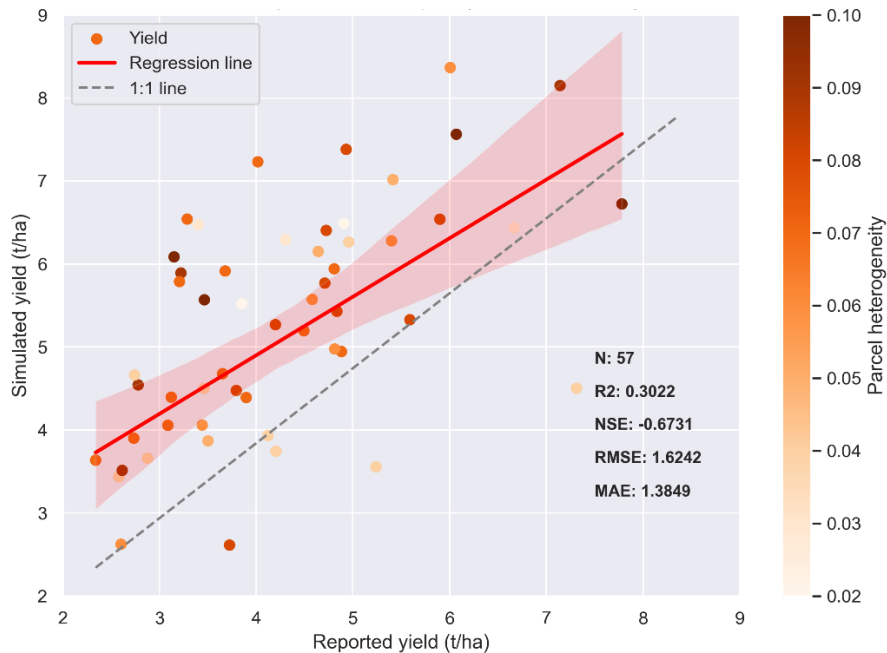


Figure 48: Model yield estimation performance after assimilation for March 27-14 April planting dates

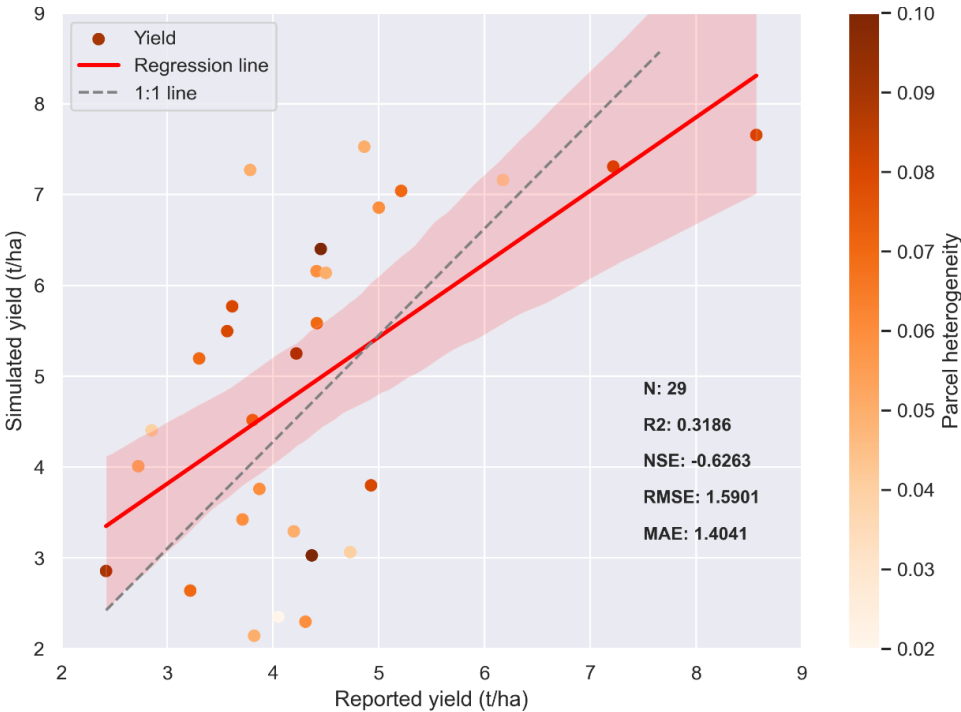


Figure 49: Model yield estimation performance after assimilation for March 27-14 April planting dates

Improving maize crop yield estimation by assimilating remotely sensed biophysical variables in AquaCrop agrometeorological model: A case study of Western Kenya

Catherine Akinyi Odera

Crop yield estimation is vital in providing insights to policymakers on the state of food security within a given country. The application of crop yield estimation modelling has been realized to provide important information for data-driven decision-making. However, the models used have some shortcomings in providing accurate yield estimations. Studies have shown that yield estimation significantly improves by assimilating remotely sensed biophysical variables into the crop estimation models to curb this challenge. This study employs AquaCrop agro-meteorological model in maize crop yield estimation, following its successful application in previous studies in arid and semi-arid areas (ASALs). The study addresses the simulation of maize yield under rain-fed conditions in Western Kenya within two counties, i.e., Uasin Gishu and Keiyo Marakwet, for 2021. Meteorological data used in the model was obtained from ERA5-land, which included precipitation, maximum temperature, minimum temperature, and potential evapotranspiration. In addition, satellite-derived FCover obtained from Sentinel-2 was assimilated into the model with the aim of improving its performance. The model simulation performance of maize yield before FCover assimilation was $R^2 = 0.198$ at the parcel level and $R^2 = 0.812$ at the ward level. After the FCover assimilation, the accuracy improved to $R^2 = 0.334$ at the parcel level and $R^2 = 0.843$ at the ward level. These results demonstrate that remotely sensed biophysical variables improves the performance of crop yield estimation models.