UCLouvain

AGRO LOUVAIN

Faculté des bioingénieurs

Improving wheat crop yield estimation by assimilation of remote sensing biophysical variable in the Simple Algorithm for Yield Estimation (SAFY) model: A case study of Spain

Auteur : Ibrahim Raufu

Promoteur(s) : Pierre Defourny

Lecteur(s) : Xavier Draye

Sophie Bontemps

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

Acknowledgements

I would like to extend my deepest appreciation to my supervisor, Prof. Pierre Defourny. Your guidance, mentorship, and relentless dedication have been instrumental in shaping the direction of my research and I am truly honored to have had the opportunity to work under your tutelage. I would also like to acknowledge Pierre Hudmont for his technical expertise and invaluable assistance throughout the project. His insights and advice have significantly contributed to the realization of this master's thesis.

To my colleagues in the GEM program, thank you for your friendship, intellectual discussions, and the sense of community we shared. Your support and friendship made the journey more enjoyable and enriched my learning experience.

I want to express my heartfelt gratitude to my wife, Mrs. Afeiye Modinat, and my newborn child, Ayman. Despite not being physically present with me during my studies abroad, their love, understanding, and unwavering support have been my constant motivation.

I am also deeply grateful to my family and friends for their encouragement and moral support throughout this undertaking.

Finally, I would like to extend my appreciation to the jury members of my thesis, Prof. Xavier Draye and Dr. Sophie Bontemps, for taking the time to read and evaluate my thesis report.

To my family and friends who impacted my life in one way or another during my study, even though you were thousands of miles away, I give thanks to you all.

Contents

List of Figures

List of Tables

Acronyms

4DVar	Four-Dimensional Variational Data Assimilation
APAR	Absorbed photosynthetic active radiation
AWiFS	Advanced Wide Field Sensor
CC	Canopy cover
CGM	Crop growth models
DAM	Dry above-ground mass
ELUE	Effective light use efficiency
EnKF	Ensemble Kalman Filter
EO	Earth observation
ERA-5	Fifth major global reanalysis produced by ECMWF
ESA	European Space Agency
ESYRCE	Encuesta sobre Superficies y Rendimientos Cultivos
ET	Evapotranspiration
FAO	Food and Agriculture Organization
fAPAR	Fraction of photosynthetically active radiation absorbed by the canopy
fCover	Fractional cover
GPR	Gaussian processes regression
HBM	Hierarchical Bayesian Method
IPAD	International Production Assessment Division
KF	Kalman Filter
KRR	Kernel Ridge Regression
LAI	Leaf area index
LUT	Lookup table

MCMC	Markov Chain Monte Carlo
NDVI	Normalized difference vegetation index
NN	Neural networks
NSO	National Statistics Office
PF	Particle filter
POD4DVar	Proper Orthogonal Decomposition into 4DVar
PSO	Particle Swarm Optimization
RS-LAI	Remote sensing leaf area index
RTM	Radiative transfer models
S2	Seentinel-2
S2-LAI	Sentinel-2 leaf area index
SAFY	Simple Algorithm for Yield estimation
SAFY-V	SAFY model with drought monitoring index
SAFY-WB	SAFY-water balance model
SAVI	Soil-adjusted vegetation index
SCE-UA	Shuffle Complex Evolution-University of Arizona
SDG	Sustainable Development Goal
Sen2-Agri	Sentinel-2 Agriculture
Sen4Stat	Sentinels for Agricultural Statistics
SLA	Specific leaf area
SVR	Support Vector Regression
TOC	Top of canopy
USDA	United State Department of Agriculture
UTM	Universal Transverse Mercator

1. INTRODUCTION

The need to address the challenges of food shortage and food insecurity in recent times calls for the prompt implementation of effective and efficient management tools. Agriculture has long been a critical component of the economies of both developed and developing nations. The Food and Agriculture Organization (FAO) reported that agriculture contributes 39.4% to the GDP, with approximately 67% of the global population engaged in agricultural-related occupations, and agricultural products accounting for 43% of all exports (FAO, 2000). These statistics highlight the significant role agriculture plays in supporting livelihoods, economies, and international trade. Wheat is a major food crop that is grown around the world and plays a significant role in ensuring global food security (Shiferaw et al., 2013). In developed countries, less than 50% of the available wheat (including imports) is used for human consumption, compared to more than 80% in developing nations (Awika, 2011). However, due to its use in producing energy and feeding livestock, the need for wheat has significantly increased over the past few decades. To attain the sustainable development goal (SDG) 2 set by the United Nations, which aims to achieve "zero hunger, food security, and improved nutrition, and promote sustainable agriculture" by 2030 (United Nations, 2015), timely and effective agricultural and crop yield estimation is fundamental to ensure the attainment of this goal.

Many nations around the world still employ conventional methods of crop yield estimation such as crop-cutting trials and household surveys. Due to its simplicity and potential precision, this approach has been around for a very long time despite being laborious, expensive, time-consuming, and labour-intensive. The method necessitated crop harvesting before yield could be assessed, making it vulnerable to errors that are difficult to spot. However, a number of approaches and technologies have been created recently to meet the demands of modern agricultural yield estimation. In this perspective, the use of technology like satellite remote sensing for crop yield estimation would offer the required assistance in ensuring food security and sustainability, especially with the projection of the global population increment from 7.8 to 9.9 billion by 2050 (PRB, 2020). Since crop yields are crucial to global agricultural development, precise yield predictions prior to harvest will provide the necessary support to make informed choices about various aspects of crop management and food supply chains (Jin et al., 2018).

The retrieval of biophysical parameters such as “Leaf area index (LAI), fractional cover (fCover), the fraction of photosynthetically active radiation absorbed by the canopy (fAPAR),

and plant chlorophyll concentration" (Silvestro et al., 2017; Dorigo et al., 2007) over the course of the entire growing season with very high geometric precision, depends directly on the temporal and spatial resolution of these remote sensing datasets (Dahms et al., 2016). Using these data in conjunction with crop growth models (CGMs) has shown to be successful in estimating crop yield (Bai et al., 2019; Lobell, 2013; De Wit et al., 2012; Casa et al., 2009; Kalpana et al., 2003).

Crop growth models are usually based on mathematical equations that describe the processes that drive crop growth and development (Ma et al, 2022). These equations are based on scientific knowledge about the factors that influence crop growth, such as temperature, light, water availability, and nutrient availability. For various crops, many CGMs with varying degrees of accuracy and resilience have been developed and implemented for yield prediction and monitoring of crop growth (Pan and Chen, 2021). Some of the prominent models are the World Food Studies (WOFOST) model (Diepen et al., 1989), Crop Environment Resource Synthesis (CERES-Wheat) model (Ritchie, 1985), Agricultural Production Systems simulator (APSIM) model (Holzworth et al., 2014), Crop-water productivity (AquaCrop) model (Steduto et al., 2009), Decision Support System for Agrotechnology Transfer (DSSAT) model (Jones et al., 2003), Multidisciplinary simulator for standard crops (STICS) model (Brisson et al., 2003), Plant simulation (SWAP) model, and Simple Algorithm for Yield (SAFY) model (Duchemin et al., 2008). To estimate crop yield, the biophysical variables derived from remotely sensed imageries can be incorporated into CGMs using a variety of data assimilation techniques such as calibration, forcing, and updating methods (Jin et al., 2018; Dorigo et al., 2007; Delécolle et al., 1992). The techniques have been implemented to enhance the estimation of crop yield both at the field scale (Gilardelli et al., 2019; Donohue et al., 2018; Silvestro et al., 2017; Mulla, 2013) and regional scales (Huang et al., 2019; Huang et al., 2015; Curnel et al., 2011). However, it has been noted that a particular focus was given to complex CGMs and much less to simple CGMs, even though they provide excellent means of simulating crop yield. Therefore, this master's thesis had the objective to utilize a simple CGM in conjunction with remote sensing data to simulate wheat yield.

The next chapter of this master's thesis will focus on reviewing previous studies related to yield estimation utilizing remote sensing technology, with a specific emphasis on wheat crop. After the literature review, the rationale and objective will be stated. Subsequently, the methodology employed and the datasets utilized for estimating the yield will be discussed in more detail. In the final two chapters, the results and their interpretation will be discussed.

2. LITERATURE REVIEW

It is vital for decision-makers to have up-to-date and reliable information on crop production and yield. However, monitoring and estimation of crop yields are usually hampered by the absence of consistent data. The ability to predict the yield of crops using remote sensing data is made possible by its synoptic and repetitive coverage (Lobell, 2013). Thus, the current idea in crop yield estimation utilizing remote sensing technology is thoroughly reviewed in this chapter. Four sections make up the chapter. The techniques for retrieving biophysical variables from satellite imageries are described in section 2.1, the techniques for remotely predicting crop yield are described in section 2.2, the techniques for incorporating data from remote sensing into crop simulation models for crop yield improvement are described in section 2.3, and a summary of the integration of remotely sensed data in the SAFY model is described in section 2.4.

2.1 Retrieval of Biophysical Variables from Satellite Imageries

Some of the remote sensing crop biophysical variables used in crop growth modelling are LAI, fCover, fAPAR), and plant chlorophyll concentration (Silvestro et al., 2017; Dorigo et al., 2007). However, among these biophysical variables, LAI which is the ratio between the green leaf area and the ground area (Cui and Kerekes, 2018) is mostly utilized by researchers to establish a connection between remote sensing and crop simulation models. Different techniques have been implemented to retrieve LAI from satellite imageries and they are empirical approaches, semi-empirical approaches, and physically based approaches (Verrelst et al., 2015).

In the empirical approaches also known as parametric regression approaches, a relationship is established between the biophysical variable to be retrieved and the spectral data using a fitting function such as linear regression (Verrelst et al., 2015). These approaches are usually parameterized using spectral vegetation indices. However, a thorough evaluation of potential band sequences, vegetation index expressions, and curve-fitting techniques are necessary before choosing a vegetation index model for the retrieval of biophysical variables from remote sensing imagery (Verrelst et al., 2015). Numerous spectral vegetation indices such as normalized difference vegetation index (NDVI) (Rouse et al., 1973) have been suggested and examined to retrieve LAI for various crops and phenological phases (Mananze et al., 2018; Rivera et al., 2014; Haboudane et al., 2004). Mananze et al. (2018) evaluated the capacity of hyperspectral and multispectral data to predict the LAI of maize crop at the field level using

the empirical method and obtained a root mean square error (RMSE) value of 0.35 between the predicted LAI and measured LAI. Even though the approach is simple to apply, it is limited by the fact that the relationship established for a crop in a particular condition cannot be transferred to another context. To overcome this kind of constraint, a wide sampling must be ensured, taking into account various crop development stages, distinct crop varieties, and various geographic locations (Mananze et al., 2018; Pôças et al., 2017).

The semi-empirical approaches also known as non-parametric approaches used information that is derived from existing data to establish a fitting function without making any assumptions about the distribution of the data or the relationships between the variables (Verrelst et al., 2015a). It is necessary to effectively exploit the data structure of regression models in order to create regression models with the best general performance. Typically, to achieve this, a flexible model must be developed that can combine various data structure features in a non-linear manner such as the machine learning algorithm (Verrelst et al., 2015). For instance, Verrelst et al. (2012) assessed the capability of four machine learning algorithms (neural networks (NN), support vector regression (SVR), kernel ridge regression (KRR), and Gaussian processes regression (GPR)) to retrieve LAI from Sentinel-2 (S2) and Sentinel-3 (S3) simulated data. They found that the GPR method gave a better result than the other retrieval methods, with an RMSE value ranging between 0.35 and 0.55. Despite the fastness and capability of capturing non-linear relationships between various parameters, the drawback of this approach is the constraint of time variation and location (Cui and Kerekes, 2018).

The physically based approaches involve the use of radiative transfer models (RTMs) to establish a physically based relationship between the biophysical variables and the spectral variation of canopy reflectance from the sensor without relying on in-situ data (Atzberger et al., 2015). RTMs are considered the most practical method for retrieving biophysical variables because of their capabilities to analyze earth observation (EO) data obtained through a variety of configurations in contrast to parametric models (Portillo, 2016). However, the models need to be inverted because of ill-posedness using an appropriate inversion method to estimate the biophysical variables (Atzberger et al., 2015; Verrelst et al., 2015, Kimes et al., 2000). Numerous techniques have been adopted by researchers to solve the RTMs ill-posedness problem by implementing strategies that rely on the Lookup table (LUT) inversion method which involves getting the RTM variables through the simulation of the spectral reflectance of the vegetation canopy (Verrelst et al., 2015a; Verrelst et al., 2015; Atzberger et al., 2015; Dorigo et al., 2009; Combal et al., 2003). In a study conducted by Delloye et al. (2018), the

LAI was retrieved from S2 spectral bands using an Artificial Neural Network (ANN). The study demonstrated a strong agreement between the estimated and measured LAI at the field scale, with an RMSE values ranging between 0.92 and 1.12 and coefficient of determination (R^2) value ranging between 0.57 and 0.84 (Delloye et al., 2018). They concluded that the accuracy of estimating LAI improves by incorporating spectral bands such as SWIR and red-edge that are less affected by atmospheric conditions but exhibit sensitivity to vegetation (Delloye et al., 2018; Li et al., 2014; Segl et al., 2012). In another study conducted by Pan et al. (2019), the performance of empirical models, LUT approaches, and NN approaches were evaluated for the retrieval of LAI from S2 data at the regional scale. The study revealed that the LUT method yielded better results compared to the empirical models and NN methods, with an RMSE value of 0.43 between the estimated and field-measured LAI. They concluded that the S2 red edge bands improve the LAI estimation accuracy in all three approaches (Pan et al., 2019). Another different approach to solving the ill-posedness problem is to apply a priori information about the variables from different sensors and spatial contexts (Verrelst et al., 2015; Combal et al., 2003). For example, Atzberger and Richter (2012) demonstrate the effectiveness of using spatial constraints for reducing the uncertainty in the RTM inversion development.

2.2 Estimation of Crop Yield

Remote sensing is concerned with the technical methods of gathering data on properties and assessing objects of interest through airborne and space-borne sensors (Lillesand et al., 2004). Unlike aerial images, which only capture a single moment in time, satellite imaging can regularly provide more frequent data collection which is important for monitoring crop growth. Estimation of crop yield can be done in two ways. The first approach involves utilizing an empirical model (sometimes called a statistical model), and the second strategy uses remotely sensed data and crop growth models for yield estimation (Peng et al., 2021). This section solely discusses indirect yield estimation methods with respect to the review done by Basso and Liu (2018) while the assimilation techniques in crop growth models will be discussed later.

2.2.1 Crop yield estimation using statistical models

Statistical or empirical models are the most basic approach for estimating crop yield and are still widely utilized today. Statistical models utilized spectral information such as vegetation indices and reflectance at various wavelengths which are derived from remote sensors (active and passive) as a variable to forecast seasonal crop yield and biomass (Basso and Liu, 2018).

Using remote sensing data or in conjunction with agro-environmental variables that are not remotely sensed, several types of spectral vegetation indices have been established to forecast the yield of different crops as presented in Basso and Liu (2018).

The fundamental goal of the statistical model approach is to create a regression between measured yield and data from remote sensing (Peng et al., 2021; Zhou et al., 2017). For example, Tuvdendorj et al. (2019) adopted a regression approach to establish a linear relationship between nine spectral vegetation indices derived from MODIS satellite data and the regional yield of spring wheat in northern Mongolia, and obtain a root mean square error that ranges between 410 kg/ha to 480 kg/ha. In addition to statistical models, other studies in recent years have tried to construct a relationship between satellite data and crop yield using machine-learning strategies (Ansarifar et al., 2021; Xu et al., 2020; Khaki and Wang, 2019). Researchers have also estimated yield by combining vegetation indices with agrometeorological data, soil data, and landscape data (Basso and Liu, 2018; Balaghi et al., 2008). Although the statistical model is simple for estimating crop yield, however, it is constrained by the fact that statistical yield-forecasting models created for one region are typically inapplicable to another region (Rembold et al., 2013) because of crop type, and geographical area influence. Another flaw of the statistical model is that it utilizes spectral vegetation indices that are usually dependent on only two bands to establish a relationship with crop yield. However, the bands that this method ignores may include extra crucial data that can be adapted to accurately estimate the yield (Baidar, 2020). Therefore, based on the peculiarities of the study areas' agricultural growth, researchers must choose the appropriate indices (Basso and Liu, 2018).

2.2.2 Crop yield estimation using crop growth models

Crop growth models are made up of several mathematical equations that describe how plants develop and grow in response to the climate (Basso and Liu, 2018). Crop simulation models' output variables include yield and crop biomass (Basso and Liu, 2018). The impacts of the interconnections between soil properties, local meteorological data, and management practices of crop yield is taken into consideration by crop simulation models (Basso and Liu, 2018). Portillo (2016) categorized crop simulation models into simple and complex simulation models.

Complex crop simulation models describe the relationships between the crop, soil, and atmosphere to simulate crop growth. Some of the popular complex crop models such as

WOFOST, CropSyst, APSIM, STIC, SWAP, CERES-Wheat, and EPIC models, have all incorporated remote sensing biophysical variables such as LAI, fCover, and fAPAR for estimation of crop yield at local and regional scale. The flaw of these models is that they require many input parameters and detailed information, which may not all be obtainable at a large scale. However, at a local scale, these models are useful for yield monitoring and simulating the impact of various management practices and environmental conditions (Portillo, 2016).

Simple crop simulation models on the other hand are easy to implement and require only a few parameters to simulate crucial physiological crop processes (Portillo, 2016; Padilla et al., 2012). These models are based on the light use efficiency (LUE) concept (Monteith, 1972) and assume that the production of biomass is relative to photosynthetically active radiation over the crop growing period (Lobell, 2013), and have the capacity to easily incorporate high-resolution remotely sensed data to change some model parameters. AquaCrop and SAFY models are examples of simple CGMs. SAFY model combines the LUE concept with the leaf partitioning function (Maas, 1993) to simulate the dynamics of green LAI, the accumulation of dry above-ground mass (DAM), and actual crop yield on a daily time step (Duchemin et al., 2008) using two (2) input variables (temperature and solar radiation required for crop growth) and fewer parameters. Remote sensing biophysical variables, particularly LAI data, can be assimilated by the model to change some of its parameters. SAFY model is not limited by geographical area and crop type as it was the case of statistical models and requires only few parameters compared to complex crop growth, hence the choice of the model in this study. SAFY (Duchemin et al., 2008) have been used with remote sensing data for estimating crop yield by numerous researchers. For example, Peng et al. (2021) used the SAFY model in combination with leaf area index (LAI) data generated from unmanned aerial vehicle (UAV) data to predict the yield of maize at the regional level under different irrigation scenario and obtained an RMSE value of 692.8 kg/ha and an R^2 value of 0.855 between the predicted and observed yield. In addition, Chahbi et al. (2014) also predicted the yield of wheat and barley at the field level using the SAFY model in conjunction with SPOT5 time-series LAI data derived through the NDVI-LAI relationship, and they found that the SAFY simulated yield were underestimated when compared to the observed yield with an R^2 value of 0.33.

2.3 Methods of Assimilating Remotely Sensed Data in Crop Growth Model

Assimilation of remotely sensed data is the process of applying remote sensing techniques to integrate canopy state variables with different information in both location and time to optimize

parameters in crop models (Jin et al., 2018). According to Delécolle et al. (1992), to properly process data assimilation for crop yield prediction, it is necessary to first differentiate between observed variables (derived from remotely sensed data), and model variables. One or more model state variables that may be obtained from satellite data can be utilized to assimilate remotely sensed data into crop models. There are several ways to incorporate biophysical data such as LAI, canopy cover, evapotranspiration, and fAPAR obtained through remote sensing, and many of them have been explored and used in various studies. However, forcing, calibration, and updating strategies are the three main strategies for incorporating remotely sensed data into crop models (Delécolle et al., 1992). The forcing strategy uses remotely sensed data to substitute the crop models' simulated data (Jin et al., 2018). The calibration method utilizes optimization strategies by minimizing the variation between the stimulated variables and observed variables to make sure that the CGM simulations are matched to the remotely sensed observations as best as possible (Pellenq & Boulet, 2004). The method of updating, also known as data assimilation, entails continuously updating the state variables of the crop model (Jin et al., 2018).

The crop growth modelers employed these approaches for crop modeling when remote sensing techniques began to provide more reliable quantifications of canopy state variables including LAI, biomass, evapotranspiration, soil moisture, and canopy cover (Jin et al., 2018). Subsections 2.21 to 2.23 include descriptions of the three main data assimilation methods.

2.3.1 Forcing approach

The forcing approach, as described by Delécolle et al. (1992), involves recreating a time series data for a crop canopy variable like LAI retrieved from satellite imagery and using these data in the crop growth model at every single simulation phase as opposed to allowing the model defining the values of that variable. Several researchers have successfully used the forcing approach to estimate crop yield. For example, Bai et al. (2019) estimated the yield of the jujube fruit tree at a field scale in western China by forcing LAI generated from Landsat 8 NDVI and a Soil-adjusted Vegetation Index (SAVI) into a calibrated WOFOST model. The study found that forced LAI assimilation improved the accuracy of the yield estimation compared to without assimilation simulation with an RMSE value of 0.74 ton/ha and 0.87 ton/ha and also coefficient of determination value of 0.62 and 0.59 in 2016 and 2017 respectively (Bai et al., 2019). `

Casa et al. (2012) applied the forcing technique to assimilate LAI data into the STIC model to estimate agronomic variables at crop management and soil properties level of knowledge. The

result shows a reasonable improvement in yield and biomass estimation. The authors pointed out that the accuracy and regularity of observed LAI may have a substantial influence on the prediction of agronomic variables, especially during the initial growth stage, and recommend an empirical method built on a calibration approach for assessing the relationship between LAI and agronomic variable. In another research conducted by Tripathy et al. (2013), LAI that was derived from Advanced Wide Field Sensor (AWiFS) NDVI data was forced into the WOFOST model for predicting the spatial yield of wheat at regional level in Punjab, India, using a correction factor. After forcing the LAI time-series data in the model, the predicted value of growth variables such as total above-ground biomass, weights of leaves, stems, and storage organs was lowered compared to before forcing, and the estimated wheat yield had an RMSE less than 0.4 ton/ha in comparison to the measured yield data.

Although the forcing method makes it straightforward to incorporate remote sensing canopy variables into the crop growth model, its primary flaw is that when observed data are incorporated into the crop model, the error is not considered, and model and observation uncertainties are ignored. Thus, the observation quality is what determines how well it performs (Hu et al., 2019). Due to this flaw, it appears that alternative assimilation procedures are preferable to the forcing strategy.

2.3.2 Calibration approach

The calibration method assumes that the crop simulation model is technically accurate but not calibrated (Delecolle et al., 1992). Thus, it is believed that the model output error results solely from the error in the chosen model parameters and initial conditions while errors from observation and model structure are ignored (Hu et al., 2019; Evensen, 2009). Cournède et al. (2012) opined that sensitivity analysis should be performed first to determine the model parameters and initial states that ought to be calibrated to produce accurate simulations for the crop model. However, running the model with a good range of model parameter values is necessary for calibrating the model uncertainty and sensitivity analysis of the crop model (Jin et al., 2018). As more observations become available and more (different) state variables are seen, more parameters can be calibrated for example LAI, canopy cover, and biomass. Numerous researchers have created and used a variety of algorithms to calibrate the optimal parameter combination for incorporating remote sensing data into various crop models. Some of the algorithms include Particle Swarm Optimization (PSO) (Jin et al., 2016; Son et al., 2016), Shuffle Complex Evolution-University of Arizona (SCE-UA) algorithm (Li et al., 2021; Wang

et al., 2014), Error minimization function (Novelli & Vuolo, 2019), Markov Chain Monte Carlo Differential Evolution Adaptive Metropolis (MCMC-DREAM) (Upreti et al., 2020, Sadegh and Vrugt, 2014) and conjugate direction optimization algorithm POWELL (Fang et al., 2011).

With the aid of an optimization technique, the calibration approach reduced the discrepancies between the crop model's generated data and data from remote sensing. It is vital to have in mind that the accessibility of well-distributed observed variables from remote sensing data, the quality of the state variables obtained from the model, the number of independent variables that need to be adjusted, and the complexity of the crop model all affect how accurate the outcomes from the calibrated crop model will be. To increase the accuracy of crop models' predictions, the calibration procedure is expected to provide more representative input crop parameters (Jin et al., 2018; Nouvellon et al., 2001). However, the key disadvantage of the method is that it necessitates more optimization repetitions, resulting in increased computing time (Jin et al., 2018)

2.3.3 Updating approach

In updating approach, the assimilation process considers both the crop simulation model and remote sensing sources of error to produce the best estimates possible (Vazifedoust et. al., 2009). The remote sensing data are adopted as external data and the state variable which is updated continuously during the assimilation process is predicted by the model (Jin et al., 2018). According to Portillo (2016), most of the assumptions made by the updating technique are that the parameters are already known, and that state variables and accompanying observations account for uncertainty. However, using updating techniques, model parameters that change over time can be estimated. During updating, these parameters are solely regarded as state variables. Also, data assimilation using updating approach gives the possibility of incorporating a combination of biophysical variables in the model.

Jin et al. (2018) presented some of the common data assimilations using updating techniques in the crop growth model and they include Kalman Filter (KF) (Vicente-Guijalba et al., 2014), Ensemble Kalman filter (EnKF) (De wit and Diepen, 2007), Particle filter (PF) (Xie et al., 2017), Four-Dimensional Variational Data Assimilation (4DVar), Hierarchical Bayesian Method (HBM), and the Proper Orthogonal Decomposition into 4DVar (POD4DVar). These assimilation techniques have been adopted by many authors in different crop models for improving crop yield estimation. For example, Vazifedoust, et. al. (2009) employed the KF strategy to assimilate time series of LAI and ET obtained from MODIS satellite images into

the SWAP crop model to forecast wheat yield at field scale and regional scale. The result obtained showed a bias of 10% (4.5 ton/ha) between the simulated yield and statistical yield data when predicting the yield at 1 month in advance with assimilation and a bias of 4 to 39% without assimilation. Wu et al. (2020) also implement the EnKF approach to improve the estimation of winter wheat yield at a field scale by integrating LAI retrieved from Sentinel-2 into the WOFOST model, and the resulting yield accuracy was improved (with an RMSE value of 512 kg/ha with assimilation of LAI and an RMSE of 818 kg/ha without assimilation) which proves the possibility of using high-resolution LAI for yield prediction at the field level. Huang et al. (2015) adopted the 4DVAR algorithm to jointly assimilate LAI and ET from MODIS data into the SWAP model to estimate winter wheat yield at field and regional levels and obtained better results in the yield accuracy with an RMSE value of 619kg/ha at the regional level when jointly integrating LAI and ET, compared to an RMSE of 889kg/ha when assimilating only LAI and an RMSE of 1561kg/ha when assimilating only ET. Li et al. (2017) also investigated the performance of PF and POD4DVAR assimilation techniques to integrate remotely sensed LAI in the CERES-Wheat model for predicting winter wheat yield at both field and regional scales in China and the result shows a better improvement in yield compared to without assimilation with POD4DVAR having better accuracy with an RMSE value of 523 kg/ha and 172 kg/ha at both scales respectively. Jiang et al. (2014) conducted a study in which the capability of the POD4DVar to integrate LAI data in the CERES-Wheat model for regional wheat yield estimation and the study's findings demonstrated that the PODVAR algorithm outperformed the EnKF method with an RMSE value of 243 kg/ha and a relative error of 2.59% at regional scale and an RMSE of 319 kg/ha at field scale.

2.4 Summary of the Integration of Remote Sensing Data into SAFY Model

The integration of remotely sensed biophysical variables in the SAFY model has allowed researchers to conduct crop yield estimation with useful results. Dong et al. (2016) used the data fusion assimilation technique to incorporate field scale LAI from MODIS and Landsat into the SAFY model to estimate the biomass of winter wheat in Eastern Canada. The outcome indicates that there is an improvement in biomass estimation due to the fusion algorithm, with an RMSE value of 1760 kg/ha and a correlation coefficient value of 0.76 showing good agreement between predicted and observed biomass. Han et al. (2021) also assimilate LAI derived from MODIS satellite data into three versions of the SAFY model (original SAFY model, SAFY-V model and SAFY-WB model) to improve the estimation of wheat yield at both field and regional scales. The authors used the shuffled complex evolution algorithm to

adjust the model parameters and the result obtained shows that the SAFY-V model performed best with an R^2 value of 0.47 compared to the SAFY-WB and the SAFY model with an R^2 value of 0.38 and 0.30 respectively. Duchemin et al. (2008), for instance, use the SAFY model in conjunction with LAI derived from Landsat Thematic Mapper data to monitor winter wheat at field scale. In the study, the yield variables were all simulated using model parameters that were calibrated using the SCEM-UA technique. When compared to field data, the outcome shows that the LAI time-series were well simulated with RMSE values ranging between 0.08 and 0.59, but the grain yield is underestimated when compared with the statistical yield on average by 0.5 ton/ha and 0.9 ton/ha when compared to the field yield and 0.48 as the correlation coefficient (Duchemin et al., 2008). Silvestro et al. (2017) also assimilate LAI that was derived from Landsat 8 satellite data into the SAFY model using the ENKF algorithm to predict wheat yield at both field and regional levels, and obtained an RMSE of 1.09 t/ha and a relative RMSE value of 18% at the field level between the predicted and observed yield. At the regional level, the predicted yield was in good agreement with the statistical yield data (Silvestro et al., 2017). Gaso et al. (2019) integrated the LAI derived from Landsat-7 and Landsat-8 images with the SAFY model to predict wheat yield at the field scale. They obtained a root-mean-squared error (RMSE) of 1532 kg/ha and an R^2 value of 0.11 between the estimated and observed yield. Manivasagam et al. (2021) also assimilate LAI data retrieved from two high resolution satellite data using a forcing approach into the SAFY model to predict the yield of wheat produced at field scale in Israel. The outcome shows that S2 and PlanetScope fused images LAI achieved higher yield prediction accuracy with an RMSE value of 690 kg/ha and an R^2 value of 0.45 compared to S2 LAI with an RMSE value of 880 kg/ha and an R^2 value of 0.35. They concluded that incorporating high resolution data from satellite imageries into crop models improves the wheat yield prediction. Ma et al. (2022) use the SAFY model to estimate wheat yield at a regional scale in Xiaotangshan Town, Beijing, by incorporating LAI from S2 satellite imagery. The authors used the SP-UCI optimization algorithm to adjust some of the model parameters and concluded that the data assimilation strategy is particularly useful for estimating yield because the estimated yield correlated properly with the observed yield data with an R^2 value of 0.49 and 0.61, and an RMSE value of 1.14 ton/ha and 1.39 ton/ha in 2014 and 2015 respectively (Ma et al., 2022). Zhang et al. (2019) also evaluate the applicability of the SAFY model to simulate winter wheat yield under different irrigation settings at the regional level through the assimilation of field measured LAI. The authors implemented the SCE-UA algorithm to minimize the error between the field observed LAI and SAFY simulated

LAI, the results shows that wheat yield was well predicted with a minimum RMSE value of 350 kg/ha obtained in the assessment of the predicted and measured yield.

The literature review revealed that only a limited number of studies have focused on field-level yield estimation due to the lack of high-resolution satellite data. Generally, high spatial and temporal resolution biophysical variables with field-sized comparability acquired throughout the growing season are required to optimize a simple crop simulation model to predict crop yield at the field scale.

The literature review also revealed a diverse range of performance outcomes, which can be attributed to the different assimilation approaches used by various authors and the limited availability of observed field data.

Moreover, it was noted that the combination of remote sensing data with CGM using updating techniques leads to significant improvement in accuracy. This can be attributed to the versatility of the technique, allowing for the assimilation of two or more variables and adjustment of model parameters based on observed field conditions.

Considering the availability of high spatial and temporal resolution data, such as S2 images, over the study area during the crop growing season, assimilating biophysical variables derived from such data with a simple crop simulation model such as SAFY might improve crop yield estimation. Therefore, the effectiveness of this approach is worth studying for improving crop yield estimation. To realize this, S2-LAI time-series data will be incorporated into the SAFY crop simulation model to simulate crop yield and the model will be validated using field measured yield data so that it can be used as a starting point for analysis of large areas of production in related locations.

3. RESEARCH RATIONALE AND OBJECTIVE

3.1 Research rationale

Based on the relevant literature covered in Chapter 2, previous studies have demonstrated the effectiveness of both complex CGMs like WOFOST and simple models like SAFY in estimating crop yield, particularly at the regional level, when combined with remote sensing data. This is mainly due to the lack of high-resolution satellite data with regular revisiting times, which sets challenges for accurate field-level yield estimation. Furthermore, complex CGMs require many parameters and extensive field datasets for yield estimation, which are challenging and time-consuming to obtain, limiting their suitability for this study.

Considering these limitations, the SAFY model is chosen for estimating wheat yield in this study, due to its simplicity and fewer parameter requirements. In addition, the availability of high-resolution S2 data through the Copernicus program of the European Space Agency (ESA) provides a valuable opportunity for field-level analysis in this study. S2 data provides high spatial, spectral, and temporal resolution observations of the Earth's surface (Vuolo et al., 2018), making it suitable for CGMs and enhancing field-level yield estimation. By utilizing S2 data, this thesis intends to overcome the limitations of conventional and labour-intensive yield estimation methods in the study area, facilitating the adoption of more efficient and cost-effective approaches. Therefore, the main goal of this master's thesis is to use S2 data to achieve the research objective stated below.

3.2 Research objective

The main objective of this research is to integrate the biophysical variable derived from remote sensing data into the SAFY agrometeorological model to improve the estimation of wheat crop yield. With a focus on improving wheat crop yield estimation, the research question of this thesis is stated below.

To what extent does the assimilation of LAI retrieve from Sentinel-2 into the SAFY model contribute to improving the accuracy of yield estimation at field level compared to the SAFY model's yield estimation without assimilation?

The interest of this research is justified by considering the following key elements: Firstly, comprehensive yield data collected through in-situ measurements creates a strong basis for understanding wheat yield dynamics in the study area, ensuring the applicability of the results. Secondly, the study includes simulation models that differentiate between rainfed and irrigated fields, enabling a more precise analysis of wheat crop yield under varying water availability conditions. Finally, through the application of the differential evolution optimization approach, the study optimize the parameters of the SAFY model, enhancing its performance and improving the estimation of wheat crop yield.

To accomplish the goal of this research, the following steps were outlined:

1. Ingestion of the SAFY model using meteorological time-series data and estimation of yield.
2. Retrieval of the leaf area index (LAI) from Sentinel-2 satellite data.
3. Assimilation of the LAI into the SAFY model using calibration method to estimate wheat yield at field scale.
4. Validation of the results of the estimated wheat yield with and without LAI assimilation using actual field measurement data.

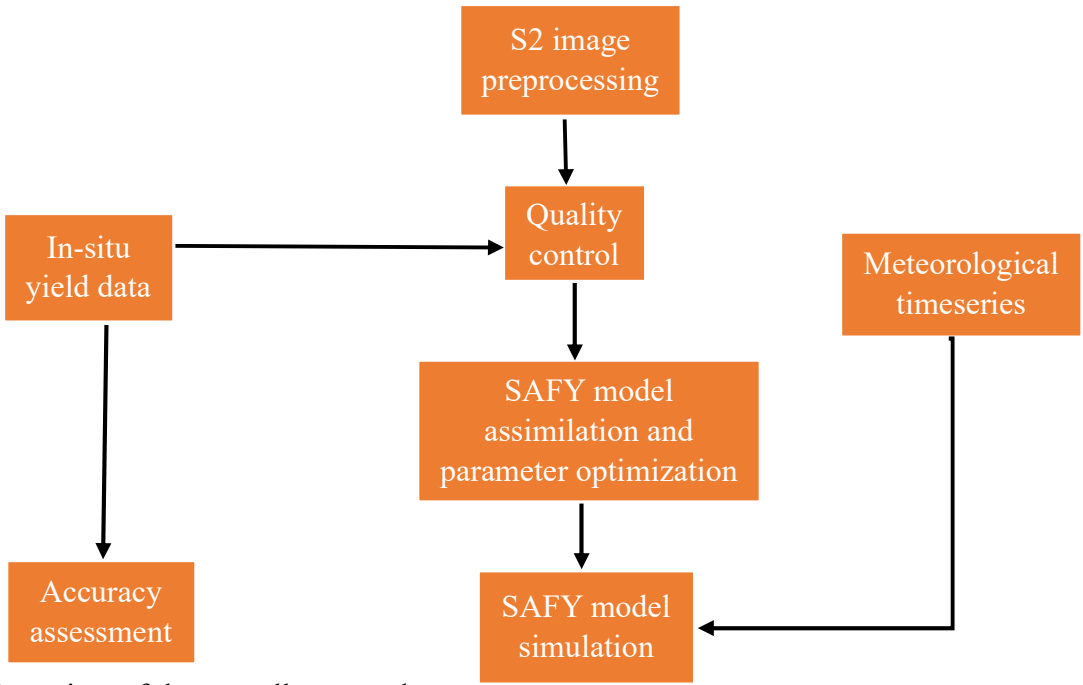


Figure 1: Overview of the overall approach

4. MATERIAL AND METHODS

This chapter provides an explanation of the methodology employed in this master's thesis. The first section of this chapter gives a background of the study area and provides a description of the datasets that were utilized for the research implementation. The second section begins by providing a summary of all the steps of this study. This is followed by a detailed explanation of the procedure for LAI estimation, a comprehensive description of the SAFY model, an explanation of the model simulation, and a step-by-step account of the calibration, validation, and performance assessment based on statistical metrics.

4.1 Study area and Datasets

This section provides background information about the research area and the datasets that have been considered to execute the research procedures.

4.1.1 Research area

The research area (Figure 2) is based on two regions, which are part of the main agricultural production areas of Spain. The regions are Castile and Leon and Castile-La Mancha. Castile and Leon is an autonomous community situated in the North-western part of Spain (41° 39’ N, 04° 45’ W) which encompasses a wide plateau with a mean altitude of 700 metres and the climate type of the region is continental, with January being the coldest month and July being the warmest month. The annual mean temperature of the region is about 13.71 °C and the annual mean precipitation is around 400-600 mm (Royo and Briceno-Felix, 2011). Some of the important crops grown in the area are wheat, barley, rye, oat, sunflower, and vineyard. Contrarily, Castile-La Mancha is situated in the South-eastern part of Spain (39° 51’ N, 04° 01’ W) and the region has a mid-latitude steppe climate type with January and July being the coldest and warmest month respectively. The annual mean temperature of the region is about 17.04 °C and the annual mean precipitation is around 450-600 mm (Royo and Briceno-Felix, 2011). Important crops grown in the area include wheat, grapes, olives, and barley.

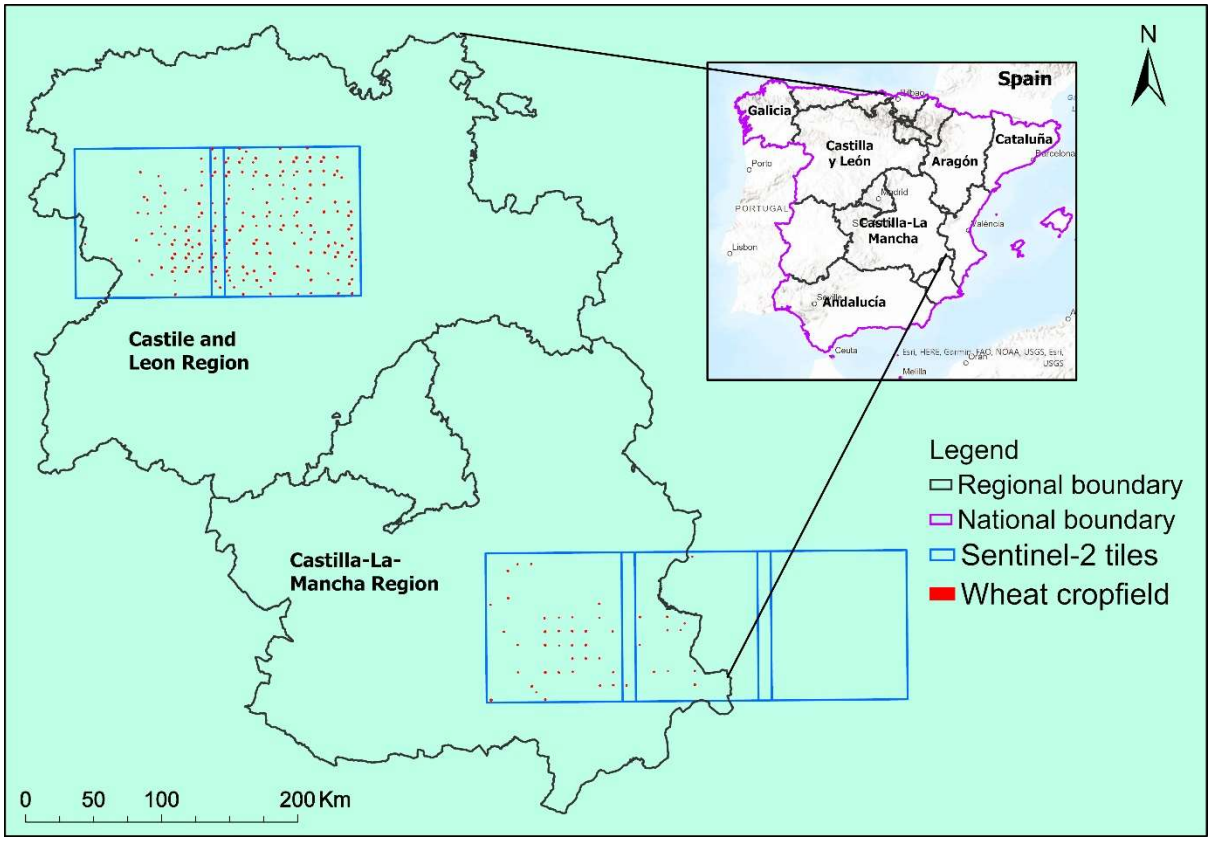


Figure 2: Research area (Esri et al., 2012)

4.1.2 Meteorological data

In this research, daily mean temperature (°C) and global solar radiation (MJ/m²) are the two meteorological datasets that are employed. The datasets are based on ERA-5 Land reanalysis which are delivered in a spatial resolution of 0.1° by 0.1° in latitude and longitude, and the data are available from 1950 to the present (Sabater, 2019). Hourly mean temperature and global radiation data from October 2018 to December 2019 were downloaded from the ERA-5 Land climate data store (https://cds.climate.copernicus.eu/) and were converted to daily data using Python code. Figure 3 and Figure 4 shows the seasonal variation in daily mean temperature and daily global solar radiation in 2019 in the two study areas. For the two areas, the highest daily mean temperature happened in July and the lowest daily mean temperature happened in January while the highest daily global solar radiation happened in June and the lowest daily global solar radiation occurred in December.

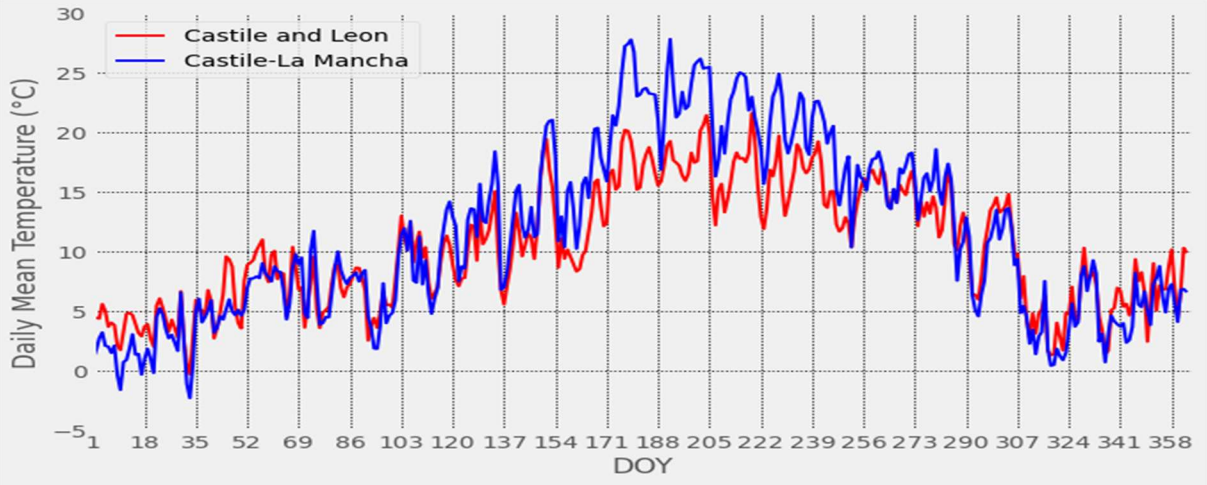


Figure 3: Daily mean temperature in 2019 in Castile and Leon, and Castile-La Mancha (https://cds.climate.copernicus.eu/)

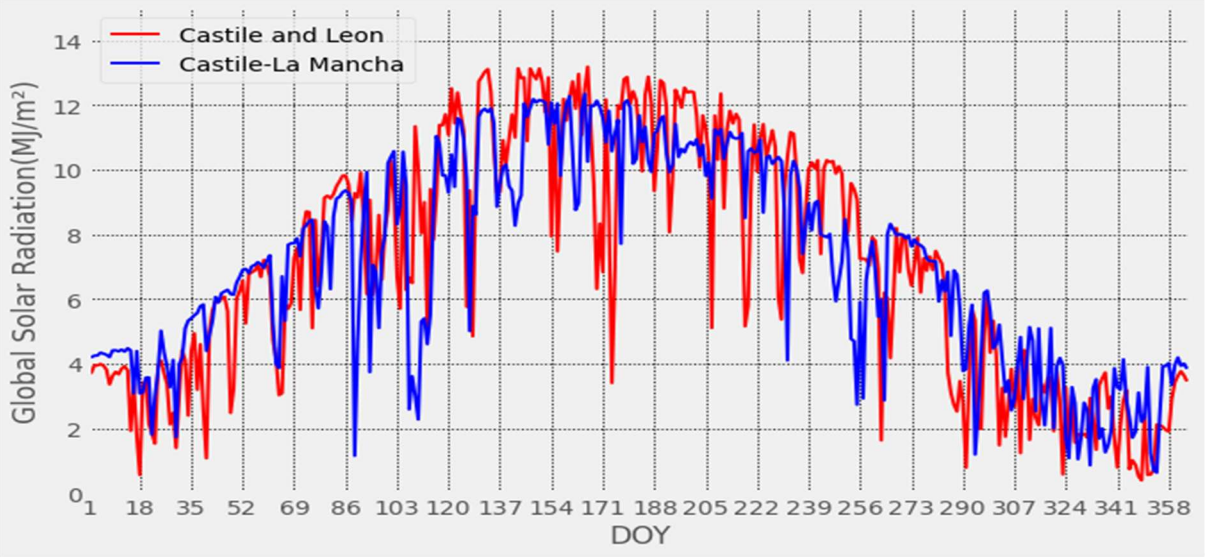


Figure 4: Daily global solar radiation in 2019 in Castile and Leon, and Castile-la Mancha (https://cds.climate.copernicus.eu/)

4.1.3 Remote sensing data

Sentinel-2 is an optical remote sensing mission that consists of two polar orbiting satellites in the sun-synchronous orbit and allow EO data to be freely available in thirteen (13) spectral bands with three different spatial resolutions (10m, 20m, and 60m) and a temporal resolution of five days revisit period (ESA, 2015). For this research, S2 images are provided in tiles through the Copernicus programme. The S2 level 2 products provide Top of Canopy (TOC) reflectance satellite data that has been radiometrically calibrated. S2 images was downloaded for the year 2019 and five (5) tiles covering the two study areas as shown in Figure 1 were used. Sen2cor atmospheric correction toolbox was used to perform the atmospheric correction

and cloud detection on the S2 images. The properties of the S2 spectral band is shown in Table 1.

Table 1: Sentinel-2 spectral bands properties (**Source:** Weiss et al., 2020; ESA, 2015)

Band Number	Band Names	Central Wavelength (nm)	Spatial Resolution	Possible Applications
1	Coastal Aerosol	443	60	Atmosphere
2	Blue	490	10	Atmosphere
3	Green	560	10	Vegetation
4	Red	665	10	Vegetation
5	Vegetation Red Edge (VNIR)	705	20	Vegetation
6	Vegetation Red Edge (VNIR)	740	20	Vegetation
7	Vegetation Red Edge (VNIR)	783	20	Vegetation
8a	Near Infra-Red (NIR)	842	10	Vegetation
8b	Vegetation Red Edge (VNIR)	865	20	Vegetation
9	Water vapour	945	60	Atmosphere
10	SWIR Cirrus	1375	60	Atmosphere
11	SWIR	1610	20	Vegetation
12	SWIR	2190	20	Vegetation

4.1.4 Wheat crop yield data

The wheat crop yield data were provided by the Spain National Statistics Office (NSO): Ministry of Agriculture, Fisheries and Feed, Department of Analysis and Agricultural Statistics (https://www.mapa.gob.es/). The Ministry provides a survey on Areas and Crop yields of Spanish crops (ESYRCE which is Encuesta sobre Superficies y Rendimientos Cultivos) at national and regional levels. ESYRCE vectorizes the cropfields within a segment of 700 square meters and attributes a crop type and a yield value for some of the crops.

For this study, yield data for 851 wheat crop fields in the two regions for the 2019 growing seasons were made available through the ESA Sentinels for Agricultural Statistics (ESA Sen4Stat) project. Out of the 851 cropfields, Castile and Leon has 721 wheat cropfields while Castile-la Mancha has 130 wheat crop fields. In this dataset, the average observed wheat yields

which comprise both rainfed and irrigated fields for Castile and Leon region range from 200 to 9,300 kg/ha and 1 to 9,000 kg/ha for Castile-la Mancha.

4.1.4.1 Wheat crop yield data filtering

To ensure the utilization of higher-quality yield data in this study, several data cleaning procedures were implemented. Firstly, any yield data with a value of 1 kg/ha was removed from the dataset. Additionally, a standard (non-spatial) statistics method, as described by Gozdowski et al. (2010), was employed to identify and address unusual yield estimates. This method involved analyzing the statistical properties of the yield data, particularly their distribution characteristics. Empirical frequency distributions were computed for the wheat yield data, and thresholds were established based on the frequency curve. Data points that fell outside of the predetermined ranges, indicating extremely low or extremely high yields, were subsequently disregarded. Figure 5 shows the histogram of the wheat yield data. Moreso, the relationship between the maximum LAI (which is an indicator of the potential leaf surface area and photosynthetic capacity of a crop) and the wheat yield data was determined. In cases where the maximum LAI value is exceptionally low, it suggests suboptimal growth and development of the crop, which can result in abnormally low yields. Typically, values larger than 6540 kg/ha and between 100 and 1020 kg/ha are treated as outliers due to their low maximum LAI. Figure 6 shows the scatterplot of wheat yield and maximum LAI of some selected cropfields in Castile and Leon to verify the outliers and the plot shows there was no significant relationship between the yield value and the maximum LAI of the cropfields that were considered as outliers. However, despite the majority of the yield values falling within the expected range, a small number of data points were identified as outliers within this range. These outliers were flagged due to their unusual yield values in relation to the size of the crop area. This approach led to the elimination of 208 (24.4%) wheat cropfields out of the 851 fields for further use in this study. Figure 7 shows the histogram of the transformed wheat yield data. However, despite the non-normal distribution of the wheat yield data and its cleaning result, the assumption is that regression analysis using the wheat yield data as validation remains valid.

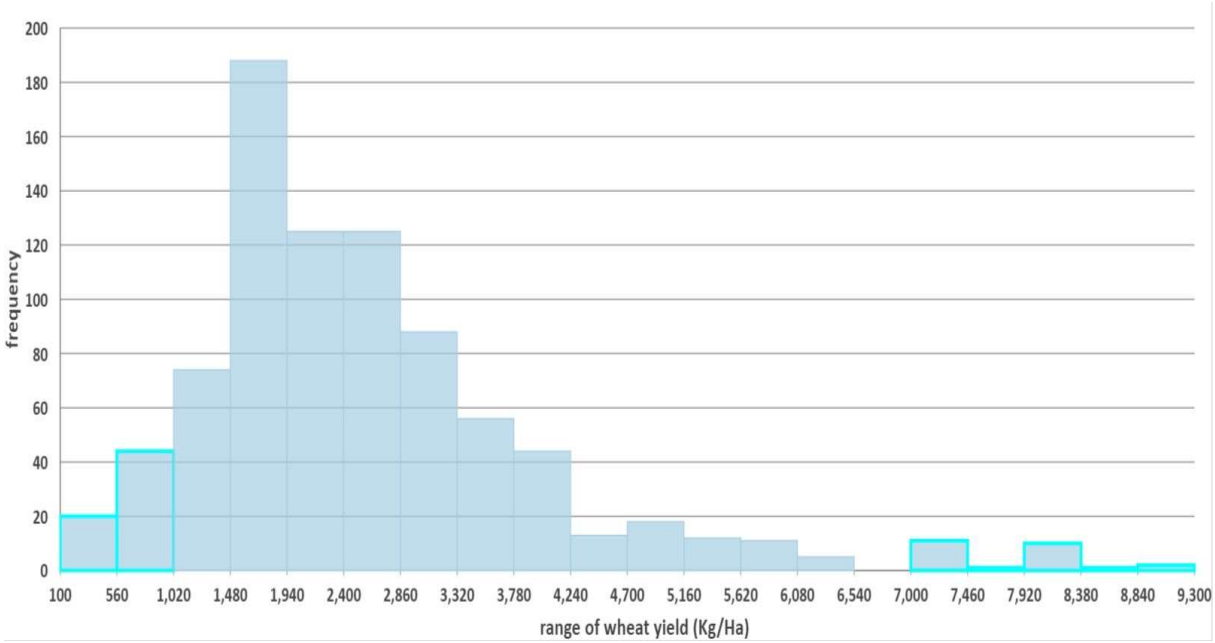


Figure 5: Histogram for wheat crop yield data (outliers are highlighted) for 2019 (n = 848).

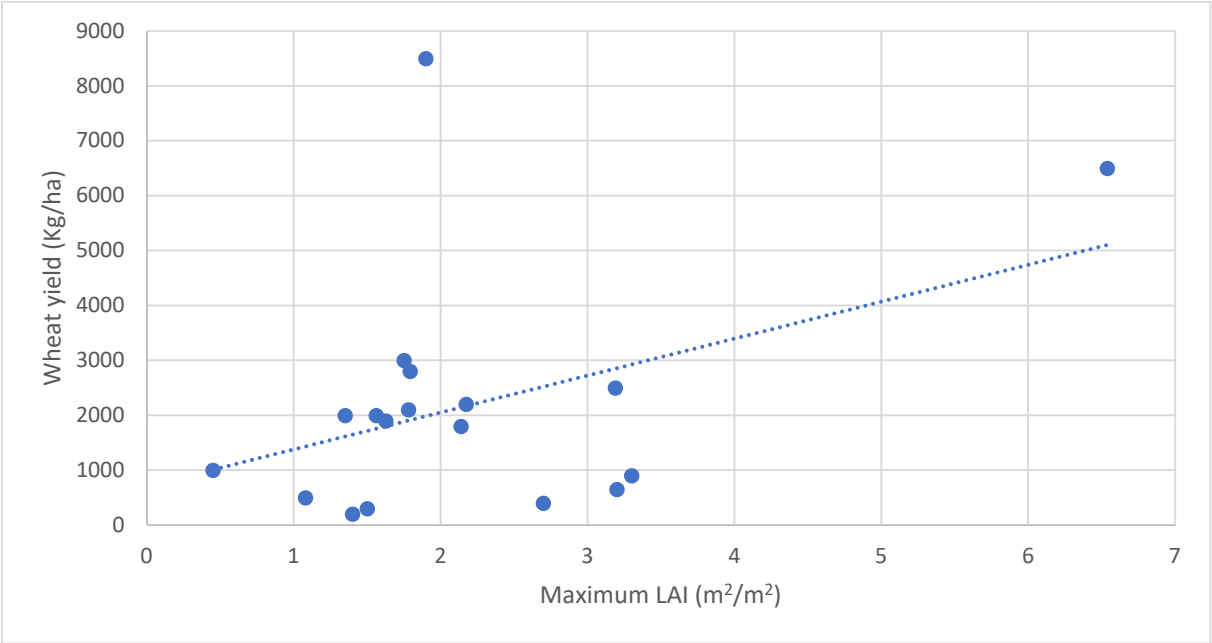


Figure 6: Scatterplot of the wheat yield of selected cropfields (in Castile and Leon) and maximum LAI (provided by the Geomatics Lab at UcLouvain).

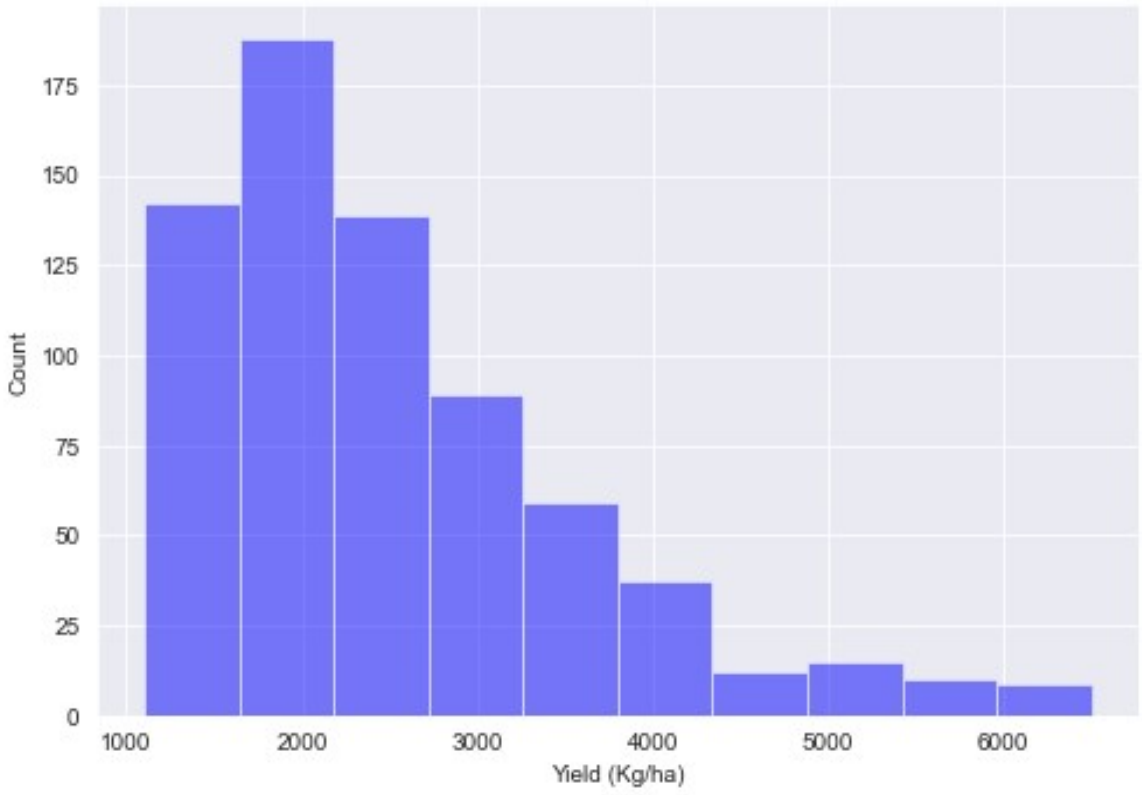


Figure 7: Histogram of the filtered wheat yield data for 2019 (n = 643).

4.2 Methods

This section provides a description of the methodology employed in this study. The research question of this master's thesis was addressed using the SAFY crop yield model and S2-LAI data. The model was initially initialized using meteorological time-series data and default wheat parameters to estimate the wheat yield. Thereafter, the model's default parameters were then calibrated by incorporating the S2-LAI. A summary of all the steps of this study is shown in Figure 8. In-depth explanations of each component are provided in the ensuing subsections.

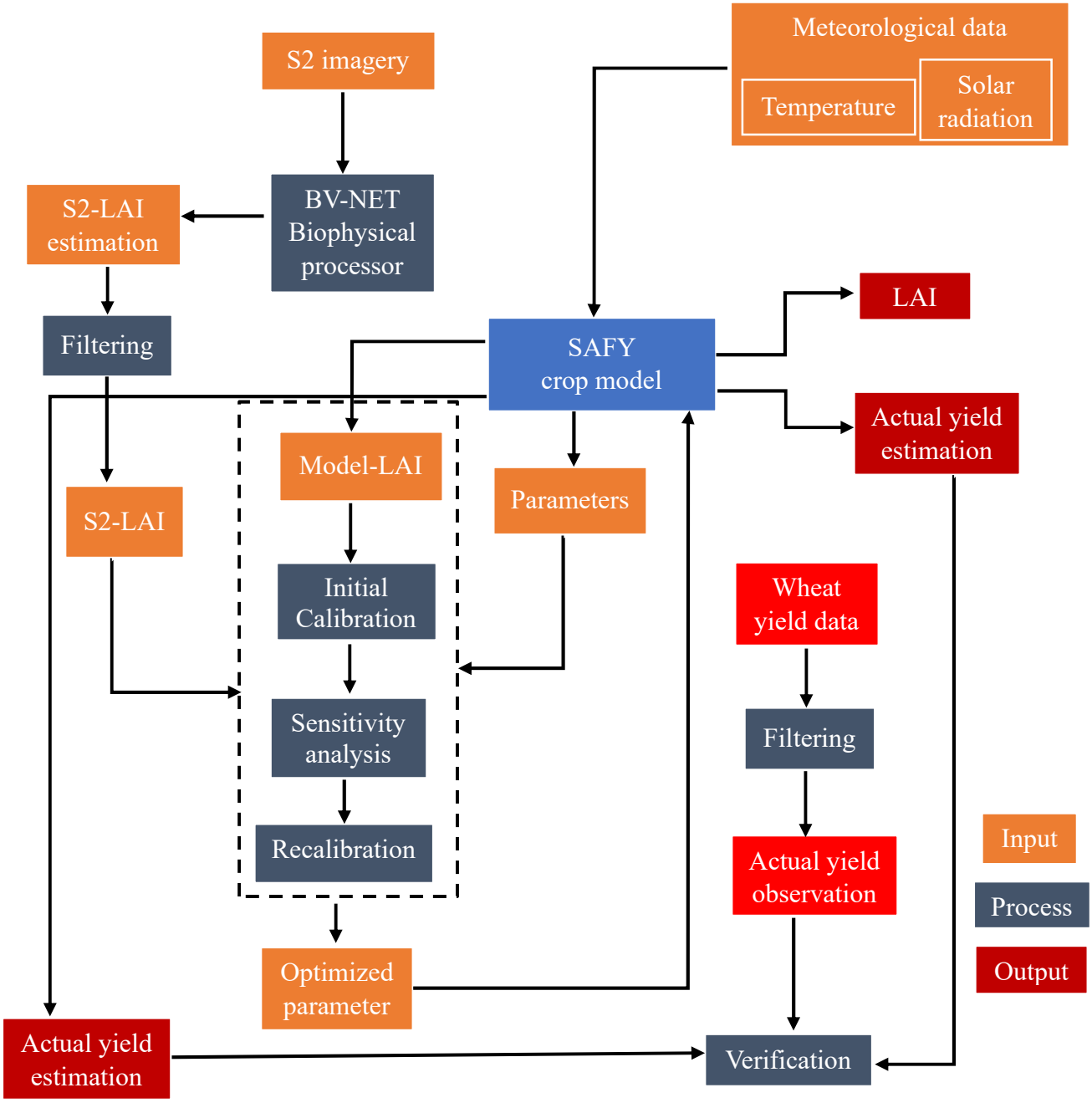


Figure 8: Flowchart of the methodology used in the study.

4.2.1 Estimation of the Leaf Area Index (LAI) from Remote Sensing Data

The total area of leaves per unit of ground area is known as the Leaf Area Index (LAI), which is expressed as a dimensionless value. It controls the exchange of energy, water, and carbon dioxide between the plant canopy and the atmosphere, as well as the amount of leaf cover in a vegetation canopy and the amount of photosynthetically active biomass (Weiss et al., 2020). When the LAI value is high, the vegetation is dense and has a high leaf cover; when the LAI value is low, the vegetation is sparse and has a low leaf cover. LAI changes dynamically over

time and is influenced by a number of variables, including species composition, phenology, and environmental conditions. LAI and reflectance have a strong nonlinear relationship. However, because of the potential for leaf distribution heterogeneity within the canopy volume, LAI cannot be directly accessed from remote sensing data apart from when utilizing directional observations (Weiss et al., 2020; Chen et al., 2005).

The LAI data used in this study was provided by the Geomatics Lab at UcLouvain. The data was retrieved from the biophysical vegetation (L3B) processor of the ESA Sentinel-2 Agriculture (Sen2-Agri) system. The L3B processor utilizes a neural network called BV-NET, which is trained to invert the PROSAIL radiative transfer model developed by INRA (Weiss et al., 2020). BV-NET provides three biophysical values related to vegetation status: NDVI, LAI, and phenology indicators. These values are derived from S2 images and are available at a spatial resolution of 10 meters in the Universal Transverse Mercator (UTM) and WGS84 projection (ESA Sen2-Agri, 2020).

4.2.1.1 LAI data screening and resampling

To ensure that better quality LAI data is used, the raw LAI time-series data was subjected to a screening and smoothing process. This process aimed to eliminate outliers and ensure that the LAI time-series followed the temporal frequency of the crop's seasonal growing cycle. In the screening step, any S2-LAI time-series data with a valid pixel value lower than 80% of the total number of pixels from which the LAI data was derived were removed across the growing stage. Next, the smoothing process involved resampling the S2-LAI timeseries to daily frequency data to ensure consistency with other datasets, and interpolating missing values using the linear interpolation technique. Finally, the interpolated data were subjected to a three-point upper envelope Savitzky-Golay function (Gu et al., 2006) to estimate the wheat phenological development. The three-point upper envelope Savitzky-Golay filter works by fitting a quadratic polynomial to three adjacent points in the LAI timeseries. The LAI value at the centre of the three-point window is then calculated using the coefficients of this polynomial. Every LAI data in the timeseries is subjected to this process again, producing a timeseries that is smoothed and follows its upper envelope (Gu et al., 2006).

It is important to note that the S2-LAI data actually represents the effective Green Area Index (GAI) of the plants. The effective GAI refers to the leaf area per unit ground area that contributes to the photosynthetic activity of the plants (Duveiller et al., 2011). The GAI provides a more accurate estimation of the actual photosynthetic capacity of the crops. By using

the S2-LAI data, which represents the effective GAI, this study ensures a more precise assessment of the crop's growth and development throughout the growing season. Figure 9 shows an example of the interpolation and smoothing result for a cropfield in the study area. In the figure, it can be seen that the smoothed LAI data representing the effective GAI, captured the crop phenology and also preserved the shape of the original S2-LAI data.

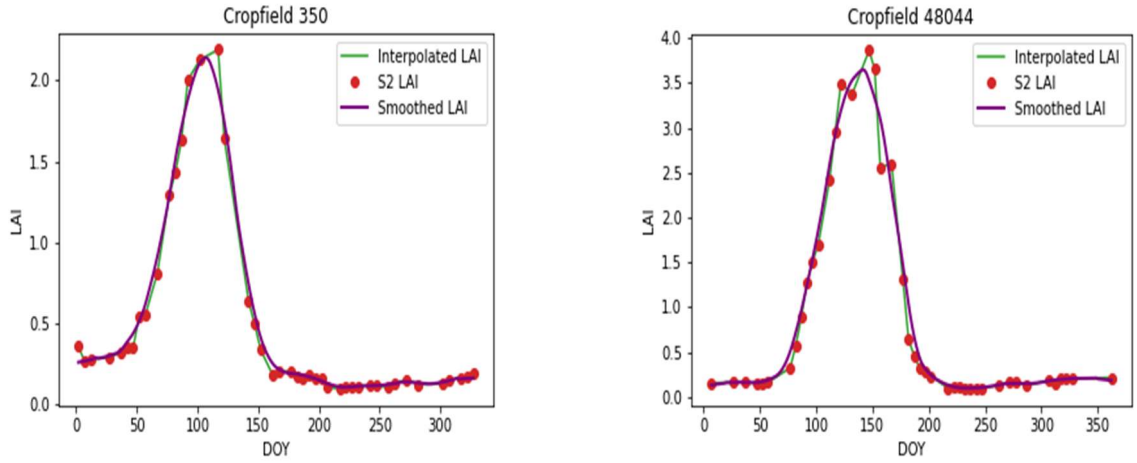


Figure 9: Example of the LAI interpolation and smoothing result for a rainfed field (left image) and irrigated field (right image).

4.2.2 Overview of the SAFY crop growth model

SAFY is a crop growth model that is based on the light use efficiency (LUE) concept (Monteith, 1977), a simple theory that connects the total dry phytomass produced and the photosynthetically active solar radiation (APAR) that plants absorb (Duchemin et al., 2008). It simulates LAI, the accumulation of dry above-ground mass (DAM), and yield on "a daily time step from the day of plant emergence (D_0) to the day of complete leaf senescence" (Duchemin et al., 2008). Leaf senescence starts when the sum of air temperature (ΣT_a) gets to a certain threshold (S_{TT}), and it increases according to the rate of senescence (Rs) factor. The leaf senescence ends when LAI has dropped to a value that is equal to zero or lower than its starting value, signifying complete senescence.

The daily average air temperature and daily incoming global radiation are required as input data to run the model with fourteen (14) parameters which can be split into three groups as shown in Table (2). The model does not require soil data, rainfall data, and management information because of the assumption that the impact of these elements are represented when the simulated LAI is fitted to the measured LAI data by adjusting some of the model parameters. A description of the SAFY model process is shown in Figure 10.

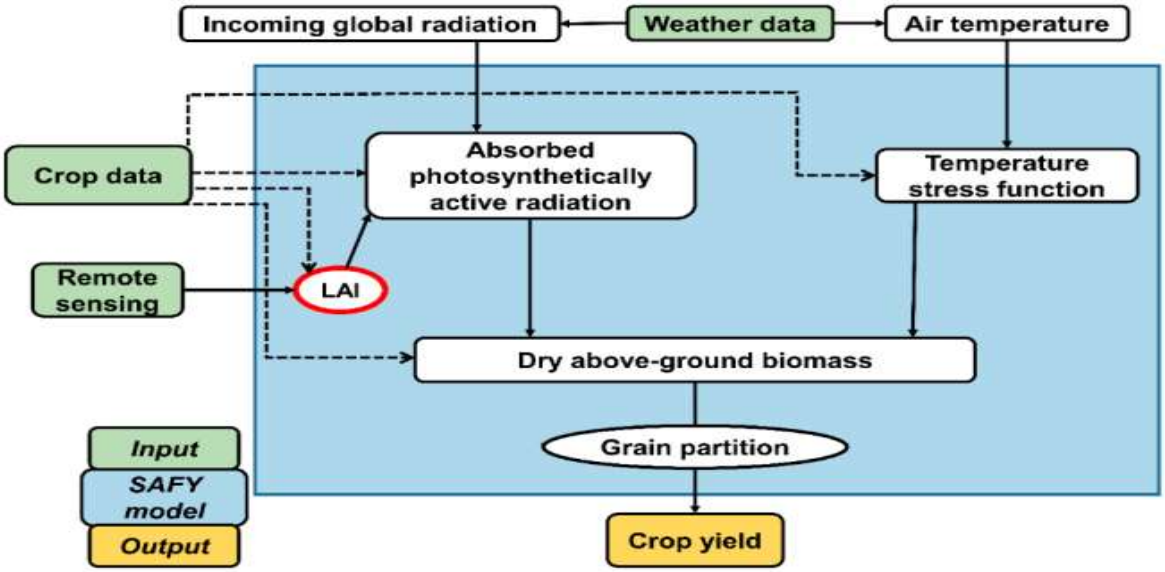


Figure 10: SAFY model flowchart (Manivasagam et al., 2021).

The DAM, LAI and yield is estimated using the following mathematical equations as described by Ma et al. (2022), Claverie et al. (2012) and Duchemin et al. (2008),

ΔDAM= APAR * ELUE * F_T (T_a) (3.1)

Where ΔDAM is the daily increase in dry above-ground mass, APAR is the absorbed photosynthetic active radiation, ELUE is the effective light-use efficiency and F_T (T_a) is the temperature stress factor.

APAR = ε_C * R_g * fAPAR (3.2)

The APAR is estimated using Equation 3.2, where ε_C is the climatic efficiency, R_g is the incoming global radiation, and fAPAR is the absorbed fraction of daily photosynthetically active radiation, which is calculated utilizing Beer's law in Equation 3.3, and k is the light interception coefficient.

fAPAR = $1 - e^{-k*\text{LAI}}$ (3.3)

The temperature stress factor F_T is estimated using Equation 3.4, where T_a is the daily average air temperature, T_{min}, T_{opt}, T_{max} are the minimum, optimum, and maximum air temperature required for crop growth.

$$F_T\,(T_a\,) = \begin{cases} 1 - \left[\frac{T_{Opt} - T_a}{T_{Opt} - T_{min}}\right]^2 & T_{min} < T_a < T_{Opt} \\ 1 - \left[\frac{T_a - T_{Opt}}{T_{max} - T_{opt}}\right]^2 & T_{max} > T_a > T_{Opt} \\ 0 & T_a < T_{min}\ or\ T_a > T_{max} \end{cases} \quad (3.4)$$

The green leaf area index is estimated from the balance between the LAI daily increase or accumulation (ΔLAI^+) and LAI daily decrease (ΔLAI^-) using Equations 3.5 and 3.7.

$$\Delta LAI^+ = \Delta DAM * Pl(\Sigma T_a) * SLA \quad (3.5)$$

Where Pl is the partition-to-leaf (which is a function that represents the leaf temperature's cumulative distribution), ΣT_a is the sum of the air temperature and SLA is the specific leaf area.

The partition-to-leaf is a function that is obtained using the sum of the air temperature in addition to two parameters (Pl_a and Pl_b) as depicted in Equation 3.6.

$$Pl(\Sigma T_a) = 1 - Pl_a * e^{Plb * \Sigma Ta} \quad (3.6)$$

$$\text{If } \Sigma T_a > S_{TT},\ \Delta LAI^- = LAI * \frac{(\Sigma Ta \quad S_{TT})}{R_S} \quad (3.7)$$

Where R_s is the rate of senescence and S_{TT} is the temperature threshold for senescence to identify when LAI starts to decline.

The crop yield is calculated by multiplying the final dry above-ground mass with the rate of grain filling factor (Py) as depicted in equation 3.8.

$$\text{Crop yield} = DAM * Py \quad (3.8)$$

4.2.2.1 SAFY model parameters

The number of parameters required by the SAFY model is fourteen (14) and they are categorized into three according to Duchemin et al. (2008) as shown in Table 2. They are crop growth parameters, phenological parameters, and crop agro-environmental condition parameters. The first category of the parameters which is based on crop growth are global to PAR incident radiation ratio (εC), extinction of radiation in canopy (k), emergence of dry above-ground mass (DAM_0), specific leaf area (SLA), minimum (Tmin), optimum (Topt), and maximum (Tmax) temperature for crop development. The second category of the parameters which are related to crop phenology are partition-to-leave function (Pla and Plb), temperature threshold to begin senescence (SenA), rate of senescence (SenB), and partition coefficient to grain (Py). The third category of the parameters which are related to the agro-environmental

condition of the crop are the day of plant emergence (D_0) and effective light use efficiency (ELUE).

4.2.3 SAFY model simulation of LAI and wheat yield

As part of the specific objective of this study to estimate yield without assimilation of the biophysical variable, an initial model simulation was conducted. Since there was no information about the agro-environmental conditions of the wheat season crop grown for each field scale and SAFY model requires the starting day and ending day of each simulation run for the cropfields, so the crop calendar provided by the International Production Assessment Division (IPAD) of the United State Department of Agriculture (USDA) (https://ipad.fas.usda.gov/rssiws/al/crop_calendar/europe.aspx) was utilized. From the USDA crop calendar, the months for the planting and harvesting of winter wheat in Spain were November and June. However, since the model requires a specific day of the year, the day to start the simulation was set to a day in October 2018 (Oct. 15) before the crop growing period and the day to stop the simulation was set to a day in July 2019 (Jul. 14) which is at the latest incidence of the end of crop growing period. Moreso, since the LAI data did not cover the year 2018, the D_0 parameter was presumed to be a day in December 2018 (Dec. 23). After all this were fixed, the default wheat parameter dataset was found in the paper of Duchemin et al. (2008) to simulate LAI and wheat yield for all the cropfields. The initial simulation was done using the meteorological time-series data (mean daily temperature and daily solar radiation). These were used as the starting point to establish the range of parameters that will be used for the free parameters that are to be calibrated in this study. However, the fundamental presumption here is that the SAFY model output will contain unambiguous systematic error as a result of inaccurate parameter values.

4.2.4 Calibration of the SAFY model parameters

In this study, seven (7) parameters were fixed according to the literature, and the remaining parameters are considered free parameters for calibration as reported in some previous studies (Ma et al., 2022; Zhao et al., 2022; Manivasagam et al., 2021; Peng et al., 2021; Dong et al., 2016; Claverie et al., 2012; Duchemin et al., 2008). These parameters were calibrated using the remote sensing data. Prior to the calibration of the SAFY model, the range of each parameter was defined based on the range found in the literature. The value and references of these parameters are shown in Table 2.

Table 2: SAFY crop growth model variables and parameters with their preliminary values

Parameter description	Notation	Unit	Initial value and range	Reference
Fixed				
Global to PAR incident radiation ratio	εC	-	0.48	Ma et al. (2022), Manivasagam et al. (2021), Duchemin et al. (2008)
Extinction of radiation in canopy	k	-	0.5	Ma et al. (2022), Duchemin et al. (2008)
Emergence dry above-ground mass	DAM_0	gm^{-2}	4.5	Zhao et al. (2022), Duchemin et al. (2008)
Specific leaf area	SLA	m^2g^{-1}	0.022	Duchemin et al. (2008)
Minimum temperature for crop development	Tmin	ºC	0	Duchemin et al. (2008)
Optimal temperature for crop development	Topt	ºC	20	Duchemin et al. (2008)
Maximum temperature for crop development	Tmax	ºC	37	Duchemin et al. (2008)
Free				
Partition-to-leave function A	Pla	-	0.01 - 0.3	Claverie et al. (2012), Duchemin et al. (2008)
Partition-to-leave function B	Plb	-	0.00001-0.01	Manivasagam et al. (2021), Claverie et al. (2012)
Temperature threshold to start senescence	SenA	ºC	200 - 2000	Duchemin et al. (2008)
Rate of senescence	SenB	ºC/day	0 - 100000	Manivasagam et al. (2021)

Day of plant emergence	D_0	day	50-150	Duchemin et al. (2008)
Effective light use efficiency	ELUE	g/MJ	1.5 - 5	Manivasagam et al. (2021), Dong et al. (2016)

4.2.4.1 SAFY model assimilation using Differential Evolution (DE) algorithm

Differential evolution (Storn & Price, 1997) is a well-known optimization approach that is utilized for real-valued functions with multiple dimensions. It employs a group of individual solutions and does not necessitate gradient information, allowing it to handle non-differentiable optimization problems. By keeping a set of candidate solutions (individuals) and merging them to generate a vector, the algorithm explores the design space (Georgioudakis & Plevris, 2020). This iterative process allows the algorithm to search for the optimal solution by iteratively improving the candidate solutions based on their performance. The individuals with the most favorable objective values are preserved for the following iteration, while those with inferior objective values are disregarded (Georgioudakis & Plevris, 2020). The process continues until a specific termination criterion is met.

In this study, two methods of DE algorithm were used to optimize the six free parameters (D_0, ELUE, SenA, SenB, Pla, and Plb) of the SAFY model using the S2-LAI data. The first method is known as the 'best1bin' approach, in which a new parameter solution is created by taking the difference between the best parameter solution in the current population and another randomly chosen parameter solution. This difference is then multiplied by a scaling factor and added to a third randomly chosen parameter solution to create the new parameter solution (Georgioudakis & Plevris, 2020). The second method is called the 'best2bin' method and it involves using two randomly chosen parameter solutions to create a new parameter solution. The difference between these two parameters is then multiplied by a scaling factor and added to the best parameter solution in the current population to create the new parameter solution (Georgioudakis & Plevris, 2020). Specifically, this algorithm was implemented in Python using the ‘differential_evolution’ function in the ‘scipy.optimize’ module.

4.2.4.2 Initial attempt to select DE algorithm

An initial attempt was made to select the DE approach to use in this study. To do these, fourteen (14) randomly selected cropfields consisting of rainfed and irrigated fields from the study area

were used to optimize all the six free parameters of the SAFY model in the first phase of the calibration using the best1bin and best2bin DE optimization algorithms. The calibration was performed by minimizing the RMSE or cost function (Equation 3.9) between the SAFY simulated LAI and the S2-LAI that has been smoothed. The algorithm starts by randomly initializing a population of parameter sets within the parameters value ranges and create a new trial vector by adding the information from multiple individuals in the population to the current parameter sets using a stochastic system. The algorithm proposes new parameter sets by randomly selecting one component from the trial vector (in the case of the best1bin method) and two components (in the case of the best2bin method) and then use the other components from the current parameter. The algorithm compares the cost function (Equation 3.9) of the new parameter set with the cost function of the current parameter set. In the end, if the new parameter set is better than the current parameter one, it makes a substitution and if it is not, the current parameter set will remain. The iterations are terminated when the cost function cannot be improved by 0.01 which is the tolerance value, or the cost function is evaluated for more than 10,000 times. The final result of the algorithm is the best optimal parameters that match the SAFY simulated LAI to the S2-LAI. Thereafter the best optimal parameters are used to simulate the LAI and wheat yield for the cropfields.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(LAI_{model,i} - LAI_{S2,i})^2}{n}} \qquad (3.9)$$

Where LAI_{model} is the LAI simulated by SAFY model and LAI_{S2} is the S2 LAI.

The result obtained for the two methods shows that the S2-LAI are always captured in all the cropfield as shown in Figure 11a for one of the cropfields. However, the best1bin method was able to minimize the cost function better than the best2bin method with an average RMSE value of 0.18 and R^2 value of 0.988 between the observed and simulated LAI across the 14 selected cropfields (both rainfed and irrigated fields) compared to the best2bin method with an average RMSE value of 0.24 and R^2 value of 0.978 as shown in Figure 11b respectively. In terms of simulated wheat yield, the best1bin method also gave a better result with an RMSE value of 1949 Kg/ha as shown in Figure 11c compared to the best2bin RMSE value of 1994 Kg/ha respectively. However, it is important to note that the optimized parameters vary across the 14 cropfields which results in the underestimation of the wheat yield except for the emergence day (D_0) which did not vary for all the cropfields. With these results, DE with the best1bin strategy is selected for the calibration of the free parameters of SAFY model in this study. This initial calibration attempt indicates that a good simulation of LAI can be obtained

from the SAFY model. It is important to note that DOY 0 represents January 1, 2019, in the LAI timeseries. The results of this initial attempt for each of the fields and the two methods are attached in appendices A and B.

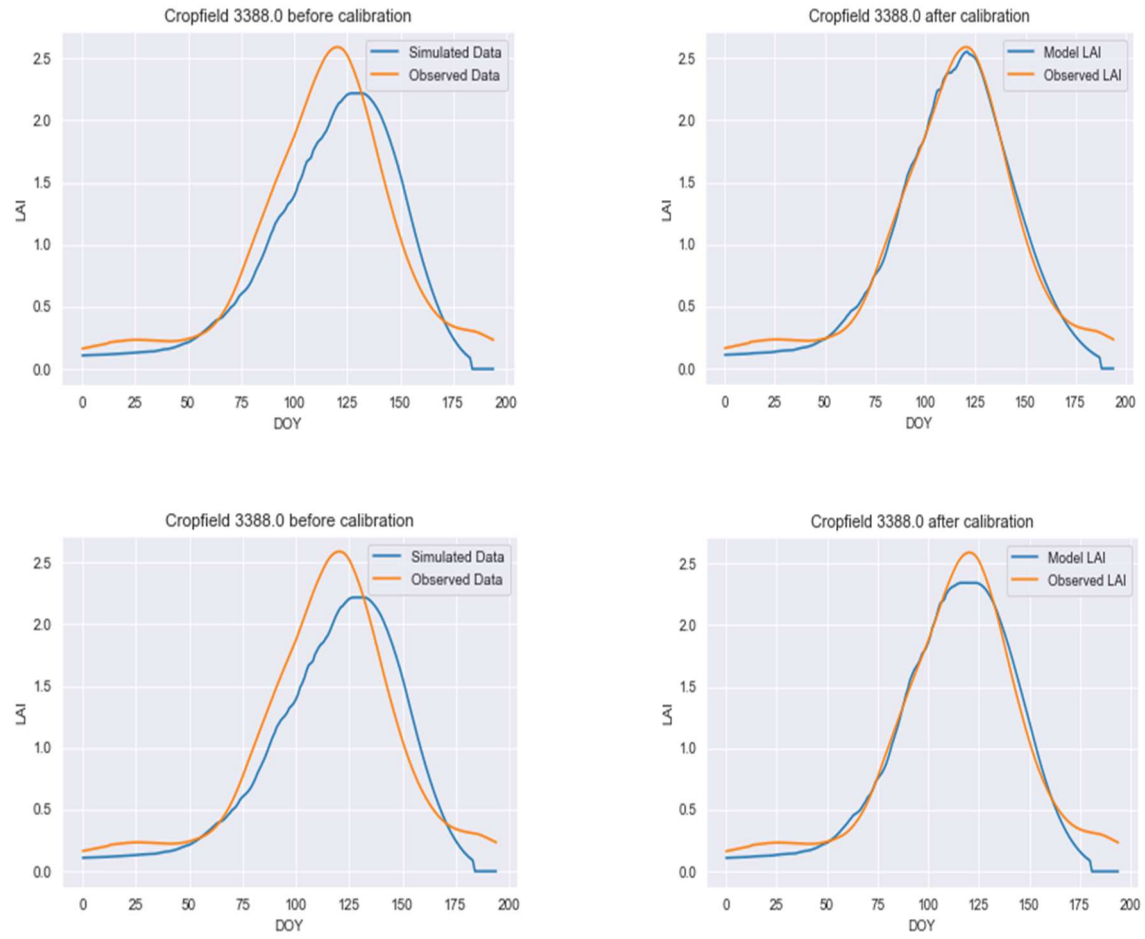


Figure 11a: Result of the DE optimization using best1bin (top image) and best2bin (bottom image).

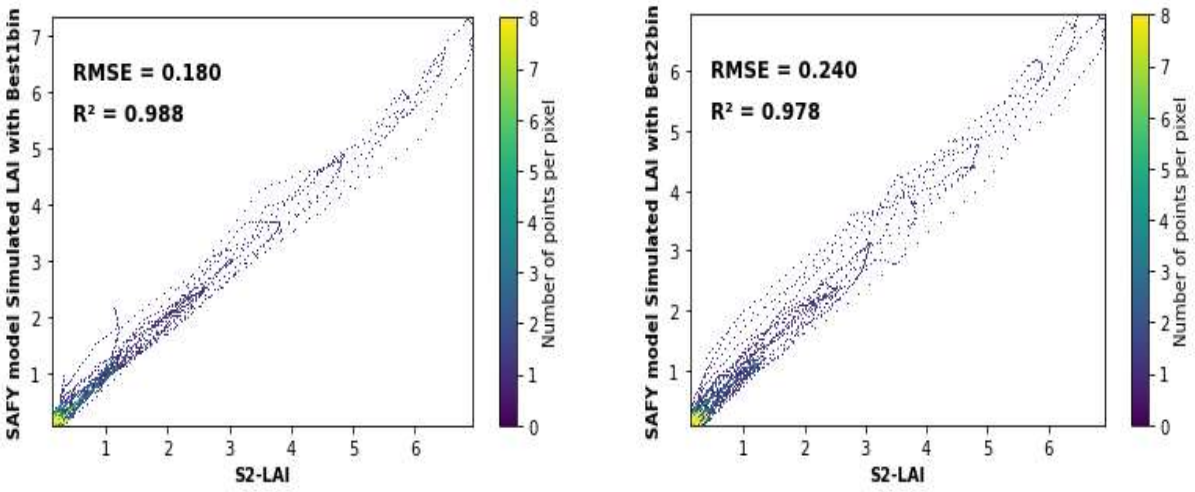


Figure 11b: Scatterplot of the comparison between S2-LAI and simulated LAI for the fourteen (14) selected cropfields using best1bin (left image) and best2bin (right image).

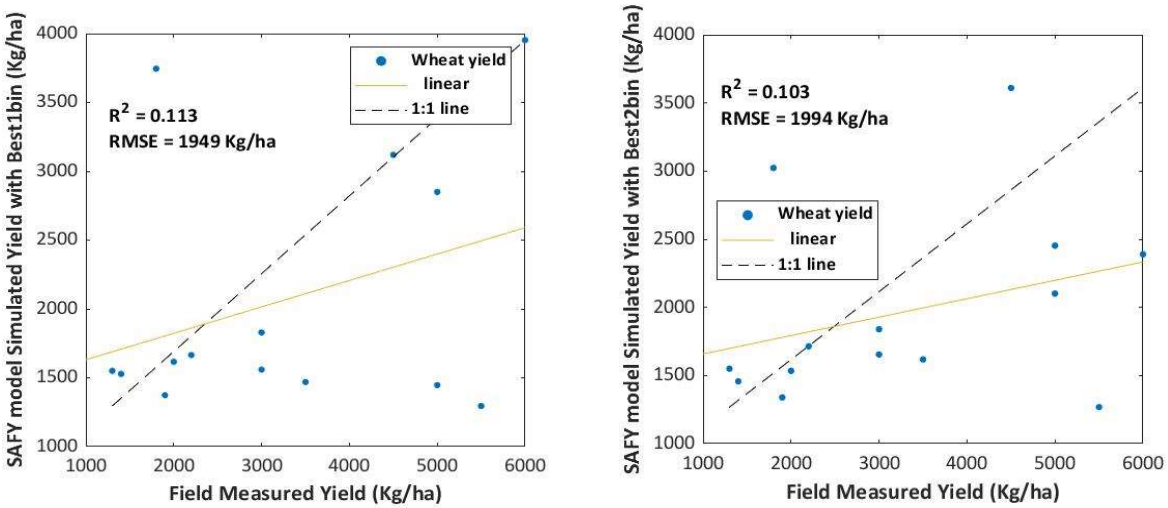


Figure 11c: Scatterplot of the comparison between observed yield and simulated yield for the fourteen (14) selected cropfields using best1bin (left image) and best2bin (right image).

4.2.4.3 Sensitivity analysis

To assess the significance of the free parameters in the SAFY model and their relationships, a sensitivity analysis was carried out using one-dimensional techniques. The comparison was made between the LAI simulated by the SAFY model and S2-LAI on 14 cropfields, with a specific emphasis on one or two of the six free parameters that affect the LAI time-series. Only one parameter was adjusted at a time during the one-dimensional sensitivity analysis, and the other five parameters were left at their optimal values for the cropfield. The target parameter was changed between -50% and +50% of its optimal value. The results of the analysis are shown in Figure 12 as differences between the SAFY model simulated LAI and S2-LAI on the 14 cropfields. The sensitivity was evaluated based on the maximum value of the root mean square error over the range of parameter variation. The results of the sensitivity analysis indicate that the effective light-use efficiency (ELUE) parameter is the most sensitive, followed by the parameter controlling the temperature threshold for the start of senescence (SenA), and the partition-to-leave function parameters (Pla and Plb). The day of plant emergence (D_0) parameter was found to be moderately sensitive, while the rate of senescence (SenB) parameter was the least sensitive.

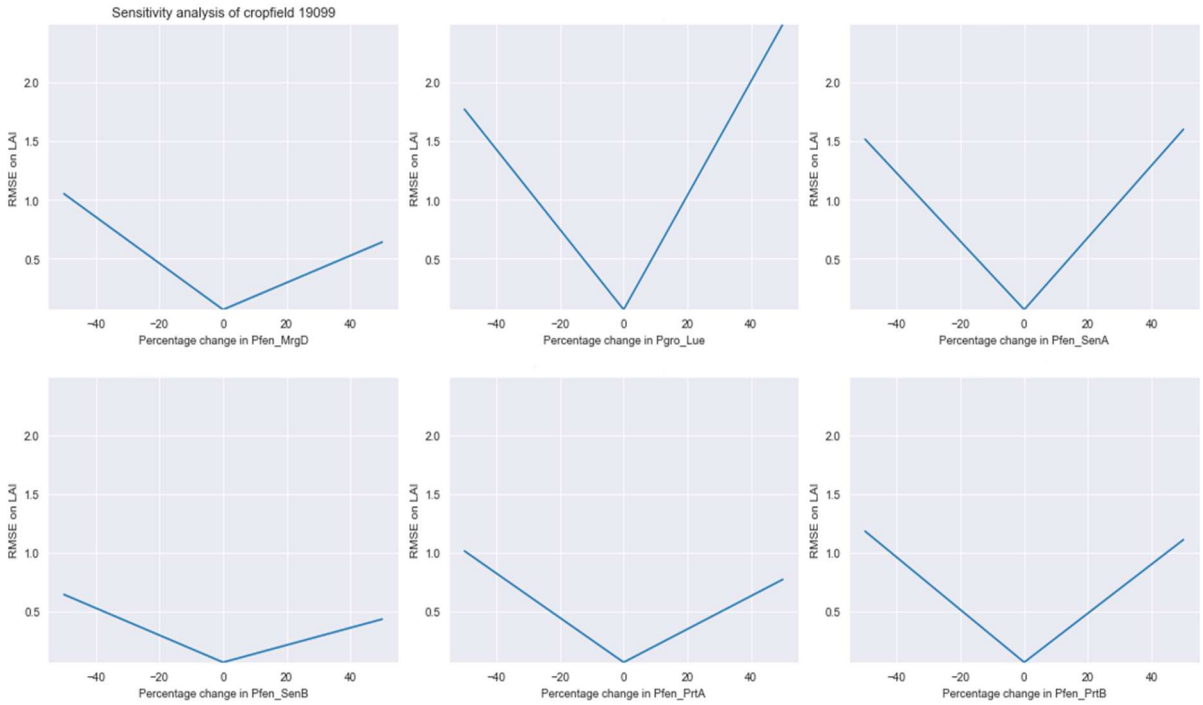


Figure 12: The outcomes of the one-dimensional sensitivity analysis for one cropfield. Each subplot corresponds to a specific targeted parameter, as indicated by the title. The plot displays the root mean square error (RMSE) between the LAI simulated by the SAFY model and the S2-LAI values for each of the six parameters that influence the LAI time-series. The parameter in each subplot is varied within the range of ±50% around its optimal value found in the initial simulation.

4.2.4.4 Day of plant emergence (D_0) trials

Since the information about the agricultural practices of the wheat crop grown in the study area was not available by cropfield, it is important to know this parameter so as to enable the model to simulate a better estimate of LAI and wheat yield. The initial calibration attempt on the 14 cropfields shows that the day of plant emergence was found constant for all the fields. With this non-variation, a test trial was conducted to determine the optimal day of plant emergence, and this was conducted by fixing the day of emergence to be 60, 65, 70, and 75 in the model while the remaining parameters were optimized. The result obtained from the trials shows that emergence day at 65 gave a better yield result with the least RMSE value of 1937 Kg/ha between the simulated and observed wheat yield compared to when the day of plant emergence is fixed at day 60 (RMSE value of 1992 Kg/ha), day 70 (RMSE value of 1963 Kg/ha), and day 75 (RMSE value of 1943 Kg/ha). Thus, the day of plant emergence (D_0) parameter is fixed at day 65 (which is equivalent to Dec. 18, 2018) in this study. The result of these trials is shown in Figure 13.

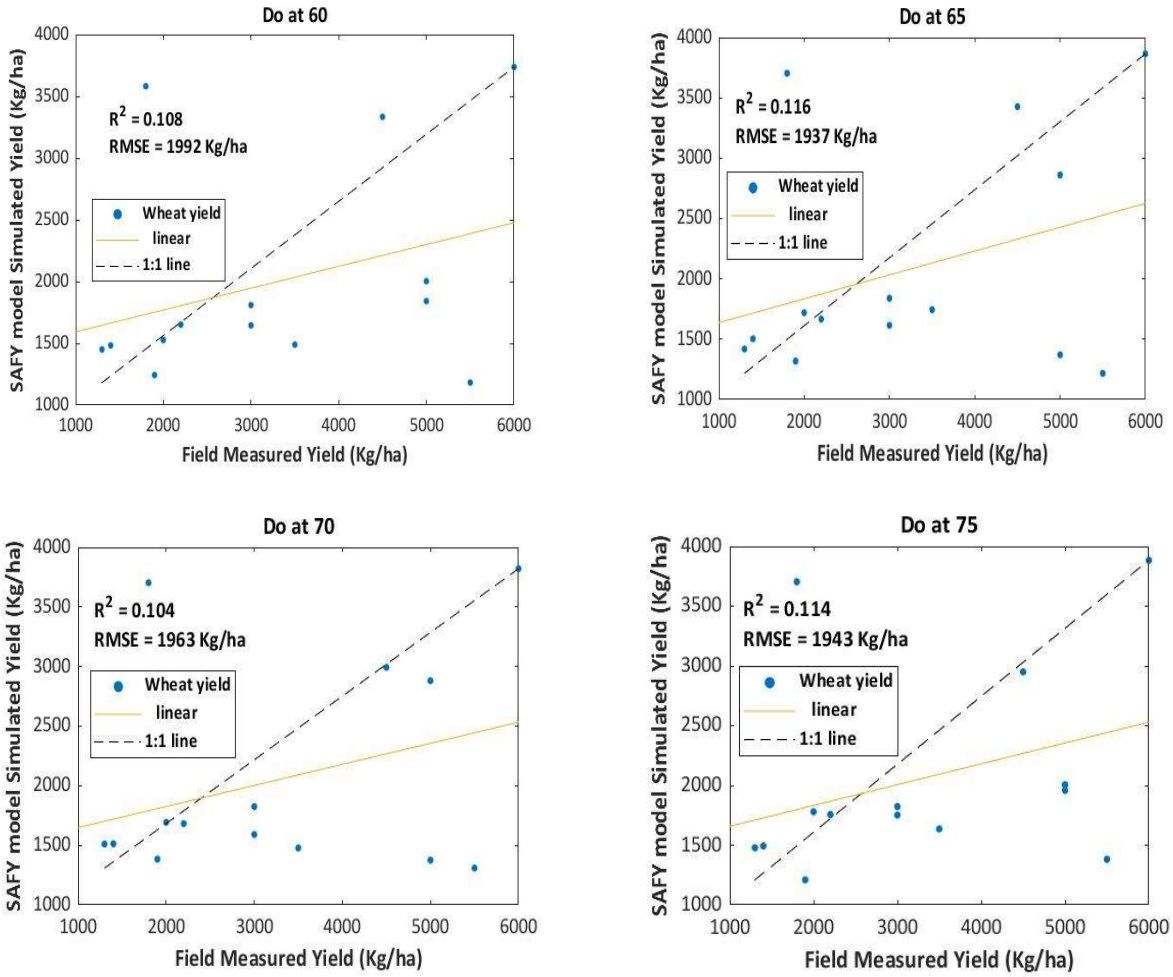


Figure 13: Variation of the day of emergence across the cropfield.

4.2.4.5 Estimation of LAI and wheat yield

With the day of plant emergence (D_0) identified and fixed for the cropfields, the remaining free parameters (ELUE, SenA, SenB, Pla, and Plb) were optimized for all the 643 cropfields (consisting of 557 rainfed cropfields and 86 irrigated cropfields) as these parameters vary field by field due to the agricultural practices and wheat varieties. The optimization was run for the rainfed and irrigated field separately with the objective of minimizing the RMSE between the LAI simulated by the model and S2-LAI to get the optimal values of the five free parameters of the SAFY model. The optimization process was the same as the one applied to the 14 cropfields (based on DE with best1bin strategy) but generalized to all the cropfields.

To obtain the wheat yield, another parameter that is important is the grain filling factor which relates the above-ground biomass to the yield when the crops are harvested. Studies indicate that the harvest index grain filling factor of wheat typically ranges from 0.45 to 0.55 when grown under normal conditions (Ma et al., 2022; Raes et al., 2009). However, the application of irrigation, nitrogen fertilizers, or the presence of diseases can either increase or decrease the

grain filling factor, depending on the timing and severity of these conditions (Ma et al, 2022). In adverse conditions, the grain filling factor can be significantly reduced to a range of 0.005-0.3, which is an important factor to consider when estimating crop yield using the optimal value (Ma et al., 2022; Han et al., 2021). So, in this study, the grain filling factor of 0.051 was used as most of the cropfield in Spain are rainfed cropfield.

4.2.4.6 Linear regression model between the wheat yield and LAI

The possibility of combining RS data with yield in order to make a model was investigated. We utilized the maximum S2-LAI values obtained for each cropfield and randomly selected half of the dataset (n = 322) to establish a linear regression model with the field-measured wheat yield, as shown in Figure 14. The results indicate a moderate correlation between the maximum S2-LAI and the field-measured wheat yield, with an R^2 value of 0.320. To validate the regression model, the remaining half of the dataset (n = 321) that are maximum LAI simulated by SAFY model was employed. Thereafter, the estimated wheat yield using the maximum LAI was compared with the estimated wheat yield by the SAFY model.

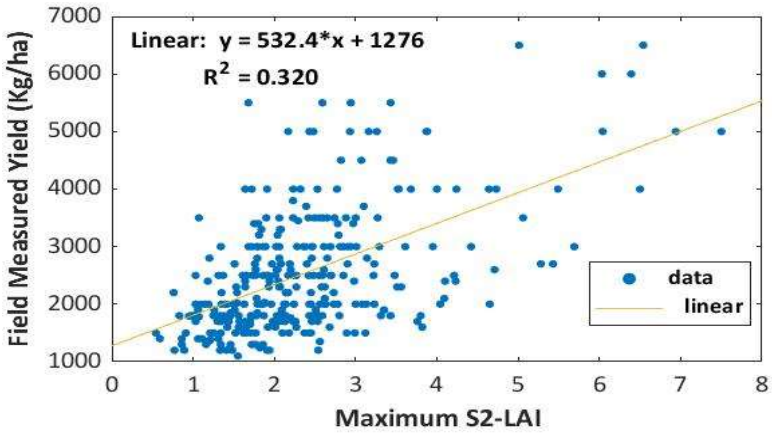


Figure 14: Relationship between field measured wheat yield and the S2-LAI

4.2.5 Validation and accuracy assessment of SAFY model

The performance of the SAFY model was assessed in relation to the field measured yield data (statistical yield data) using two statistical metrics which are the RMSE and coefficient of R^2. The RMSE is a statistical metric that provides a quantitative estimate of the model's output difference from observation by evaluating the root square of the average of the difference between predicted values (Y_{Model}) and measured values (Y_{Obs}) (Novelli & Vuolo, 2019) as shown in Equation 3.10. In this study, Y_{Model} is the wheat yield predicted by the SAFY model, Y_{Obs} is the field-measured yield, and n is the total number of cropfields. If the RMSE value is small, the model did a good job of predicting wheat yield; if the error is large, the model did

poorly. In this study, the RMSE is calculated in units of kg/ha (kilograms per hectare) because the field-measured yield data, is provided in the same units. The R^2 (Equation 3.11) of linear regression is another statistical metric that expresses the proportion of variation in the measured data that is explained by the variation in the predicted data (Novelli & Vuolo, 2019). The R^2 value is between 0 and 1, with a value of 1 indicating a strong linear relationship between the predicted variable and observed variable while a value of 0 indicates no linear relationship between the predicted and observed variables. The predicted variable can be considered a good fit for the observed variable if the R^2 value is close to 1. Nevertheless, it should also be noted that a good linear relationship can also be found between the predicted and observed variables at low R^2 values based on the context of the study.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_{Model} - Y_{Obs,})^2}{n}} \quad (3.10)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_{Obs,i} - Y_{Model,i})^2}{\sum_{i=1}^{n}(Y_{Obs,i} - \overline{Y}_{Obs,i})^2} \quad (3.11)$$

Furthermore, the F-statistics in Equation 3.12 was used to determine the statistical significance of the SAFY model performance results for yield estimation with and without assimilation of RS data. This assessment is based on the R^2 values obtained from the three model simulation cases. Then, the p-value for the calculated F is determined using a F-distribution calculator (https://www.statology.org/f-distribution-calculator/). The hypothesis testing is stated below.

H_0 = The accuracy of yield estimation with the assimilation of RS data in the SAFY model is equal to the accuracy of yield estimation without the assimilation of RS data in the SAFY model

H_1 = The accuracy of yield estimation with the assimilation of RS data in the SAFY model significantly improves compared to the accuracy of yield estimation without the assimilation of RS data in the SAFY model.

The decision rule is that, if the calculated p-value is less than 0.05 significant level (95% confidence level), we reject the H_0 and accept the H_1

$$F = \frac{R_a^2 - R_b^2/p - q}{(1 - R_a^2/(n-(p+1))} \quad (3.12)$$

Where R_a^2 is the R^2 value obtained with the assimilation of RS data, R_b^2 is the R^2 value obtained without the assimilation of RS data, p and q is the number of independent variables in the model, and n is the number of observations.

5. RESULTS AND DISCUSSION

This chapter gives the SAFY model's performance assessment results, obtained from the methods described in Chapter 4. Three major sections, each reporting on a different aspect of the model's performance, make up the chapter. The initial simulation of the LAI and wheat yield using the SAFY model is covered in the first section without the use of remote sensing biophysical variables. This offers a starting point for evaluating the model's performance after incorporating remote sensing data. The SAFY model's simulation of LAI and wheat yield with the integration of remote sensing biophysical variables is the main topic of the second section. The results of the model's performance after the addition of this new data are presented in this section. The chapter ends with a thorough explanation of the SAFY model's accuracy assessment and a comparison of its performance with prior research in the field, then the empirical model analysis. This offers a thorough assessment of the model's functionality and potential for use in various research fields.

5.1 SAFY model simulation without assimilation of biophysical variable

Figure 15 shows a sample LAI simulation phenology for a rainfed and irrigated cropfield using the meteorological time-series data and default parameter in the SAFY model. Comparison with the S2-LAI time-series shows that the model was not able to simulate the dynamics of wheat LAI as the pattern does not conform with each other. Figure 16 and 17 shows the SAFY model predicted yield without assimilation for rainfed, irrigated, and a combination of both cropfields. The R^2 values indicate that there is no correlation between the observed and simulated wheat yield as the model could not be able to predict the yield with respect to reality. The RMSE value obtained is 1325 kg/ha, 3289 kg/ha, and 1722 kg/ha for rainfed, irrigated, and a combination of both types respectively. These findings point to a significant discrepancy between the measured and predicted yields, which can be explained by the SAFY model's misrepresentation of leaf phenology in reality (based on the premise that S2-LAI represents reality). As a result, the predicted yield does not agree with the crop's measured yield which prompted the need to improve the model estimation with the assimilation of remote sensing biophysical variables.

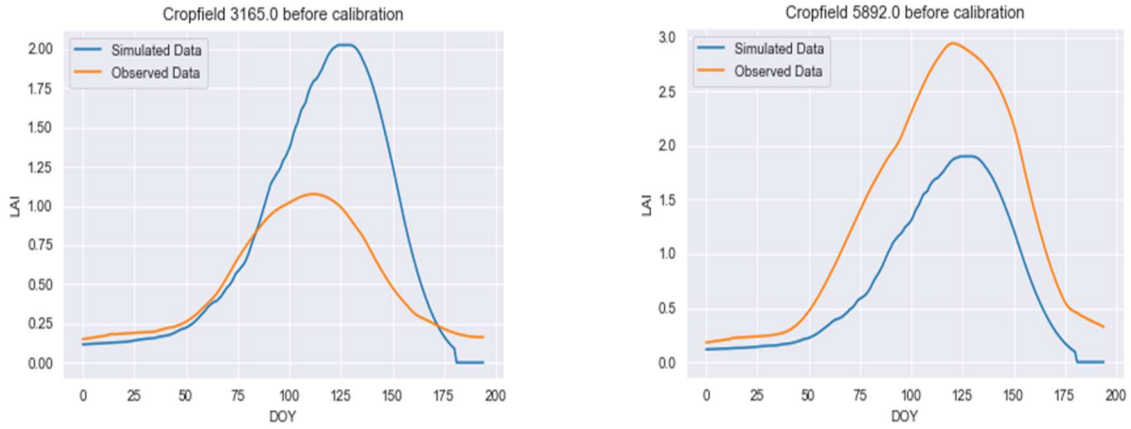


Figure 15: Sample LAI simulation without calibration for a rainfed cropfield (image on the left) and an irrigated cropfield (image on the right).

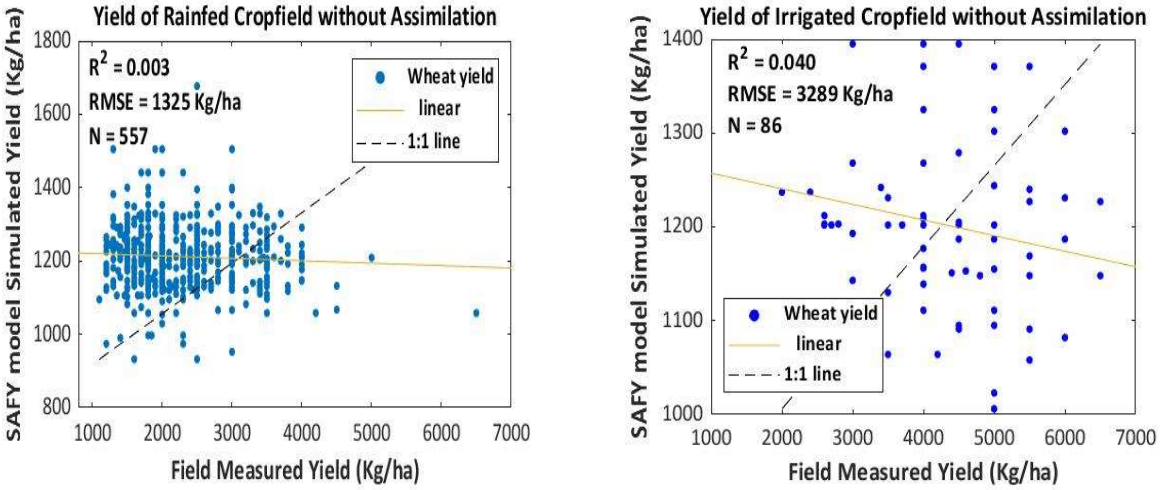


Figure 16: Scatterplot of the comparison between field measured and simulated wheat yield without assimilation of S2-LAI for rainfed cropfield (image on the left) and irrigated cropfield (image on the right).

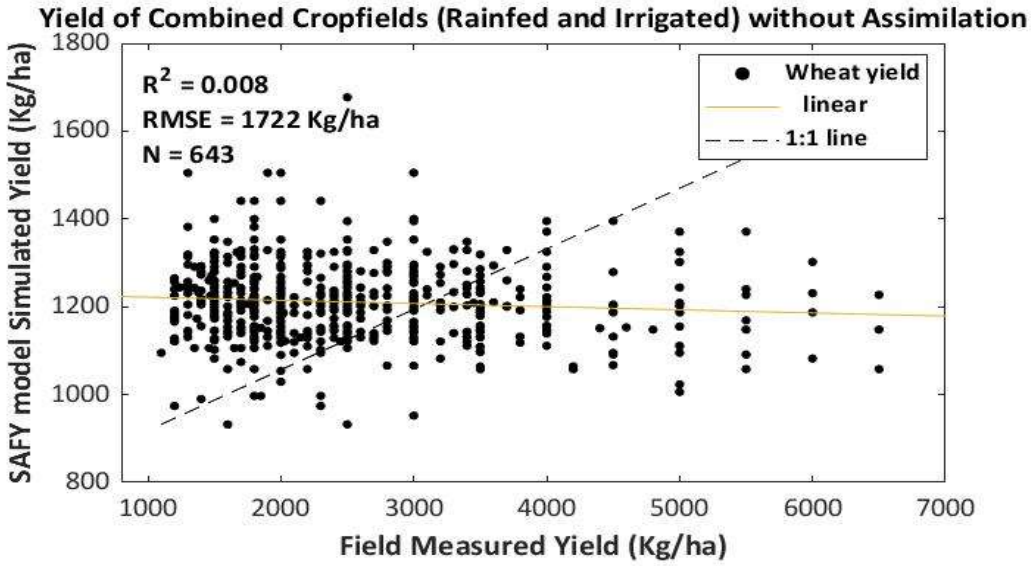


Figure 17: Scatterplot of the comparison between field measured and simulated yield without assimilation of S2-LAI for all the cropfield.

5.2 SAFY model simulation with the assimilation of S2-LAI

Figure 18 shows the LAI simulation for a rainfed cropfield and irrigated cropfield after the calibration of the model free parameters with S2-LAI. Comparison with the S2-LAI (which is used as the observed LAI) shows that the model was able to simulate the wheat crop phenology as indicated by the S2-LAI. In Figures 19 and 20, scatter plots of the simulated LAI and RS-LAI derived from S2 imagery for 2019 were generated for rainfed, irrigated, and a combination of both rainfed and irrigated wheat cropfield respectively. The R^2 value obtained for the rainfed, irrigated, and combination of both rainfed and irrigated wheat cropfields are 0.956, 0.982, and 0.967, and the RMSE value obtained are 0.197, 0.222, and 0.201 respectively. These findings demonstrate a good correlation between observed and simulated LAI values suggesting that integrating remote sensing data into the SAFY model improves the accuracy of estimating wheat LAI trajectory. This indicates the capacity of our approach to adjust the SAFY model parameters at the field level to represent the effect of the local growing conditions.

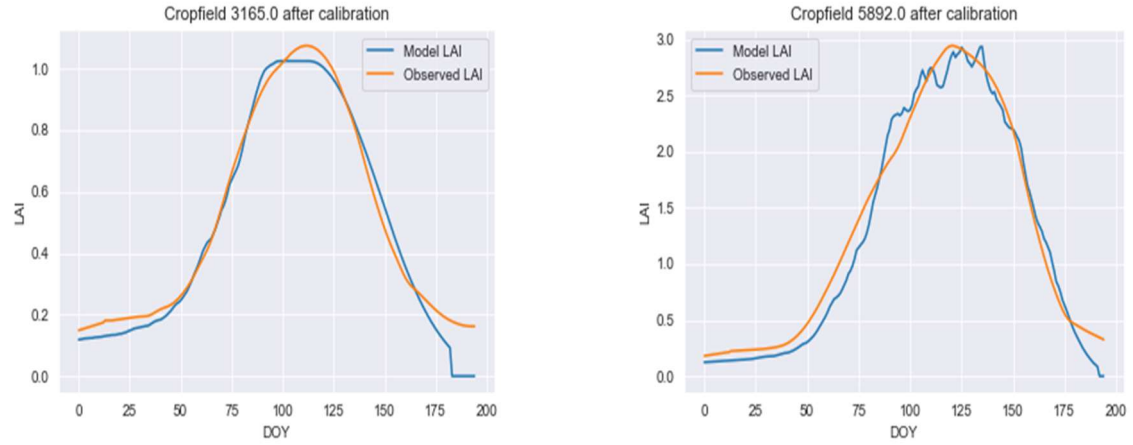


Figure 18: Sample LAI simulation after calibration for a rainfed cropfield (image on the left) and an irrigated cropfield (image on the right).

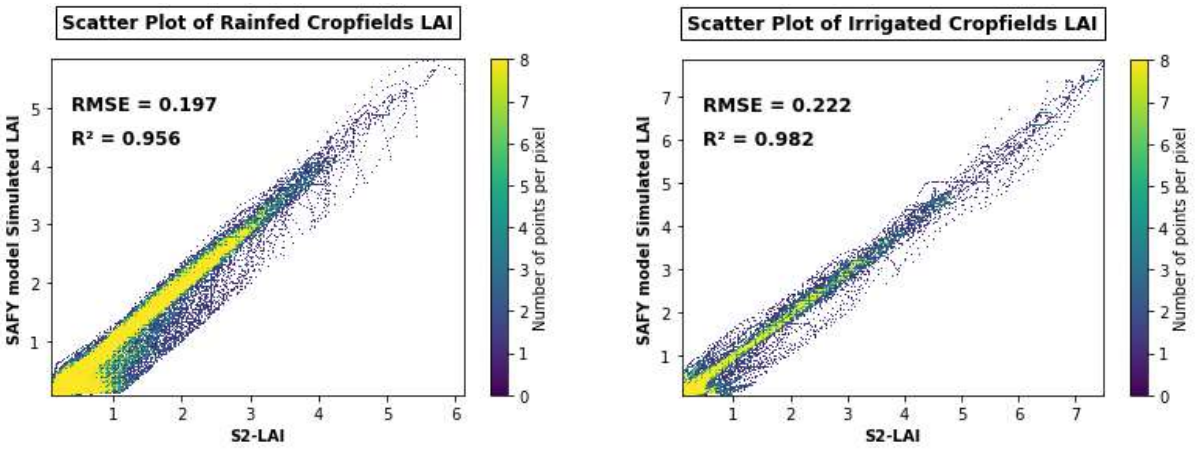


Figure 19: SAFY simulated LAI for rainfed cropfield (image on the left) and irrigated cropfield (image on the right) after assimilation.

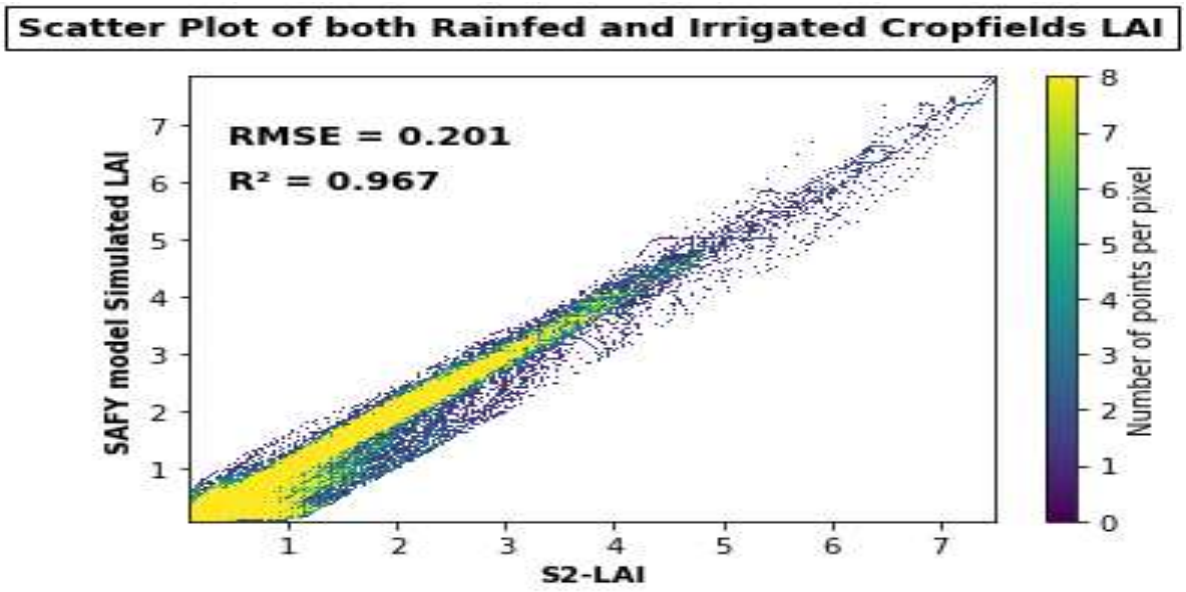


Figure 20: SAFY simulated LAI for all the wheat cropfield after assimilation.

Figure 21 and 22 illustrates the scatter plots of the comparison between the measured and predicted wheat yield for rainfed, irrigated, and a combination of both, after the assimilation of S2-LAI. The estimated R^2 values are 0.292, 0.152, and 0.378, for rainfed, irrigated, and a combination of both cropfield respectively. The RMSE obtained are 744 kg/ha, 1786 kg/ha, and 952 kg/ha. The results obtained show that the model estimation improves significantly after the assimilation of S2-LAI with the estimated yield having a moderate correlation with the observed yield likewise the huge discrepancy observed between the observed and simulated yield without S2-LAI assimilation has reduced. The R^2 values indicate statistical significance, and the RMSE values are halved with the assimilation of S2-LAI compared to without assimilation.

However, it is evident from Figures 21 and 22 that the SAFY model tends to overestimate wheat yields in cropfields with lower yields and underestimates yields in cropfields with higher yields, respectively. These limitations arise due to the simplicity of the yield estimation model used, which does not adequately account for the impact of environmental stress on crop yields. As a result, it lacks sensitivity in capturing the effects of agro-environmental stress on winter wheat yields, leading to overestimation or underestimation of the actual yields (Han et al., 2021; Chahbi et al., 2014).

Figure 21 also clearly illustrates the superior performance of the SAFY model in rainfed cropfields compared to irrigated cropfields. This can be attributed to the fact that crop growth in rainfed conditions is more sensitive to water stress than in irrigated conditions. As a result,

the SAFY model with S2-LAI shows better accuracy in simulating the development stages of wheat in rainfed cropfields, leading to more precise estimates of wheat yield. Additionally, the SAFY model accuracy in predicting wheat yield in irrigated cropfields is low as the water availability effect well captured by the S2-LAI does not take into account the specific drivers that contribute to the yield differences. The model's ability to accurately represent the distinctive dynamics and productivity potential associated with irrigated conditions is hampered by the absence of irrigation-related factors in its calculations. Figure 22 also displays the overall performance of the SAFY model in simulating wheat yield, combining rainfed and irrigated cropfields. It can be seen that the two estimates seem to be moderately correlated and most of the cropfields with low observed yield value are closer to the 1:1 line while the cropfields with high observed yield value are far from the 1:1 line. The average estimated yield of 2122 kg/ha exhibited a reasonable deviation from the observed yield of 2571 kg/ha. Although the estimated yield displayed a limited range of deviation compared to the observed yield, the SAFY model successfully depicted the lower observed yield values, primarily observed in rainfed cropfields. However, the model tended to underestimate the observed yield values that were higher, mostly observed in irrigated cropfields.

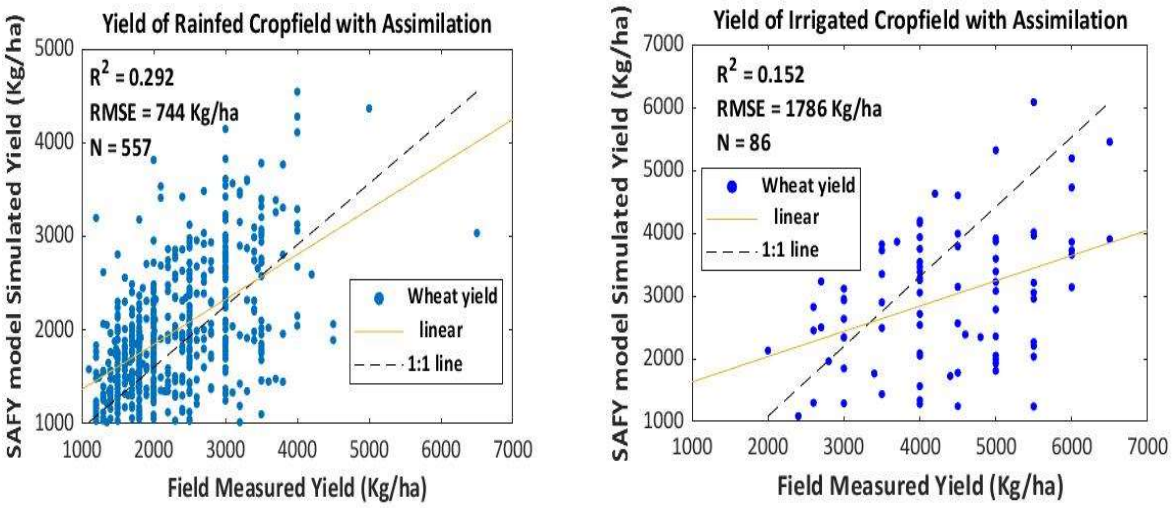


Figure 21: Scatterplot of the comparison between field measured and simulated wheat yield with assimilation of S2-LAI for rainfed cropfield (image on the left) and irrigated cropfield (image on the right).

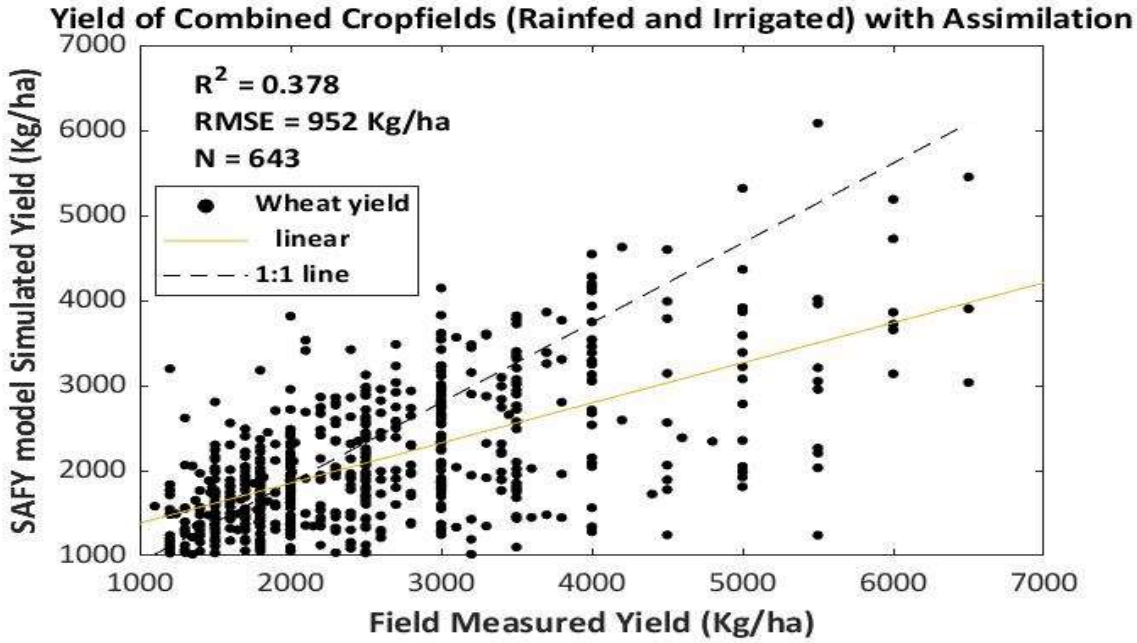


Figure 22: Scatterplot of the comparison between field measured and simulated yield with assimilation of S2-LAI for all the cropfield.

Figures 23 to 25 illustrate the scatterplot of the relationship between the SAFY model predicted wheat yield and the maximum value of S2-LAI for the cropfield with and without assimilation of LAI. It can be seen that the estimated yield and maximum S2-LAI were moderately correlated for the rainfed, irrigated, and combination of both cropfields with R^2 values of 0.245, 0.279, and 0.36 respectively with the assimilation of LAI (images on the right) while there is no correlation between the estimated yield and maximum S2-LAI for the rainfed, irrigated, and combination of both cropfields with R^2 values of 0.007, 0.068, 0.016 respectively without the assimilation of LAI (images on the left).

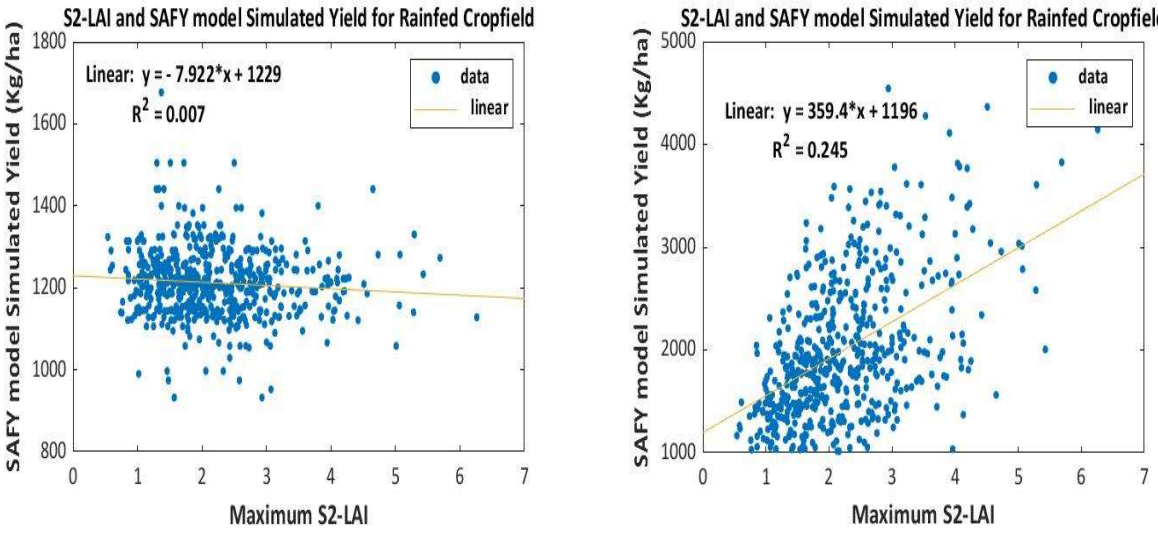


Figure 23: Scatterplot of the relationship between simulated yield and maximum S2-LAI with (image on the right) and without (image on left) assimilation for rainfed cropfield.

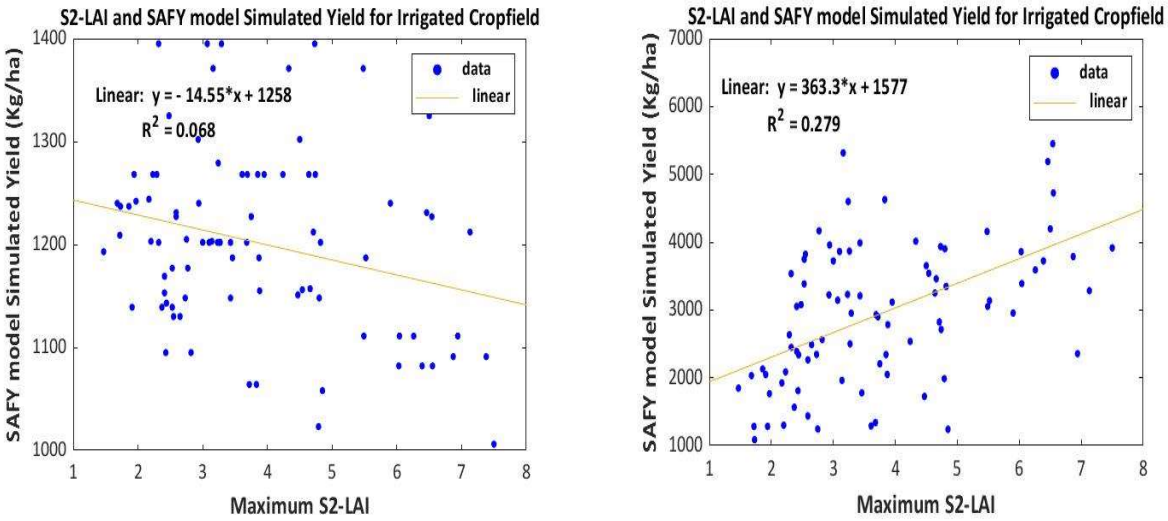


Figure 24: Scatterplot of the relationship between simulated yield and maximum S2-LAI with (image on the right) and without (image on left) assimilation for irrigated cropfield.

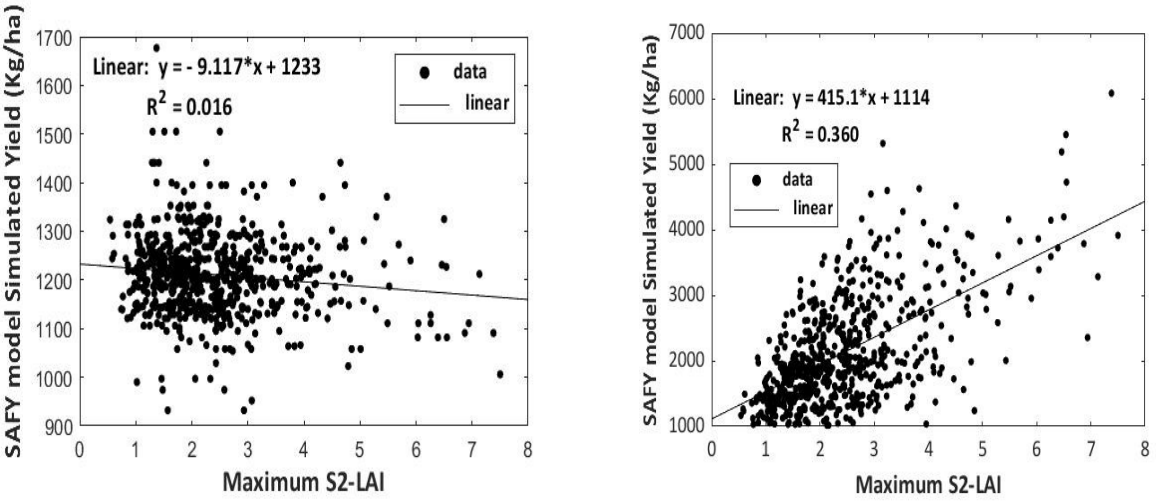


Figure 25: Scatterplot of the relationship between simulated yield and maximum S2-LAI with (image on the right) and without (image on left) assimilation for all wheat cropfield.

Figure 26 shows the relationship between the field-measured wheat yield and the estimated wheat yield using S2-LAI. The results indicate a moderate correlation between the two yields, with R^2 values of 0.321 and 0.411 for the calibration and validation cropfields, respectively. Moreso, the RMSE values are 869 kg/ha and 795 kg/ha respectively. It can be observed that the RMSE value for the validation cropfields is slightly lower than that of the calibration fields. This difference could be attributed to the fact that cropfields with maximum high LAI values simulated by the SAFY model are more in the validation fields than the calibration fields.

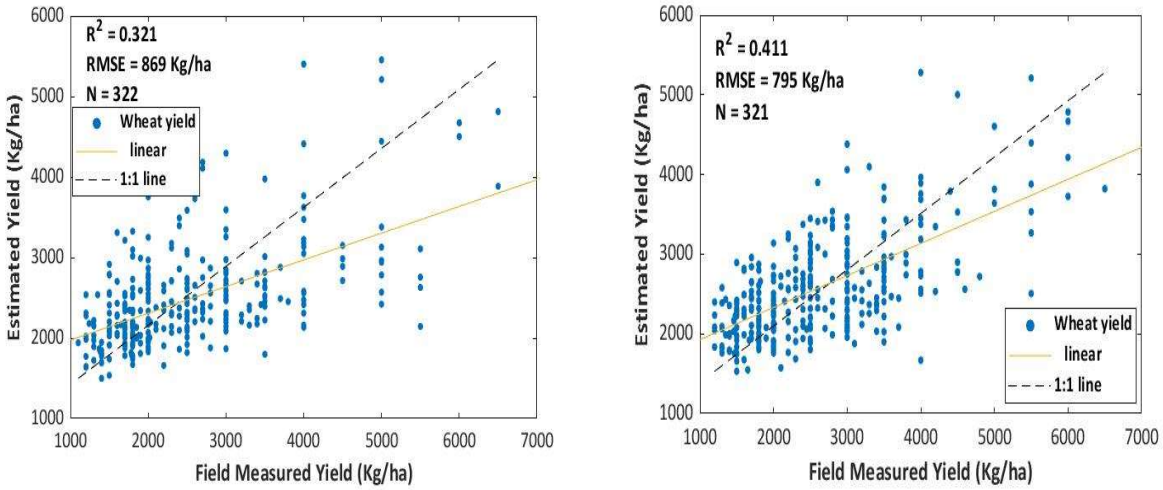


Figure 26: Scatterplot of the relationship between the field measured yield and estimated yield on calibration field (image on left) and validation field (image on right) using linear regression analysis.

Figure 27 presents a comparison between the yield estimated using the S2-LAI relationship with the field-measured yield and the yield estimated through the assimilation of S2-LAI in the SAFY model. The results indicate a moderate correlation between the two estimations, with an R^2 value of 0.380 and an RMSE value of 796 Kg/ha. However, it is evident that there exists an imbalance between the estimated yield by the SAFY model and the estimated yield by the linear regression model on certain cropfields. This discrepancy can be attributed to the yield data that was used in making the regression model which is not of high quality and also the variations in agricultural conditions within each cropfields. In addition, the limitation of the SAFY model to underestimate high wheat yield values and overestimate low yield values also contributed to the difference.

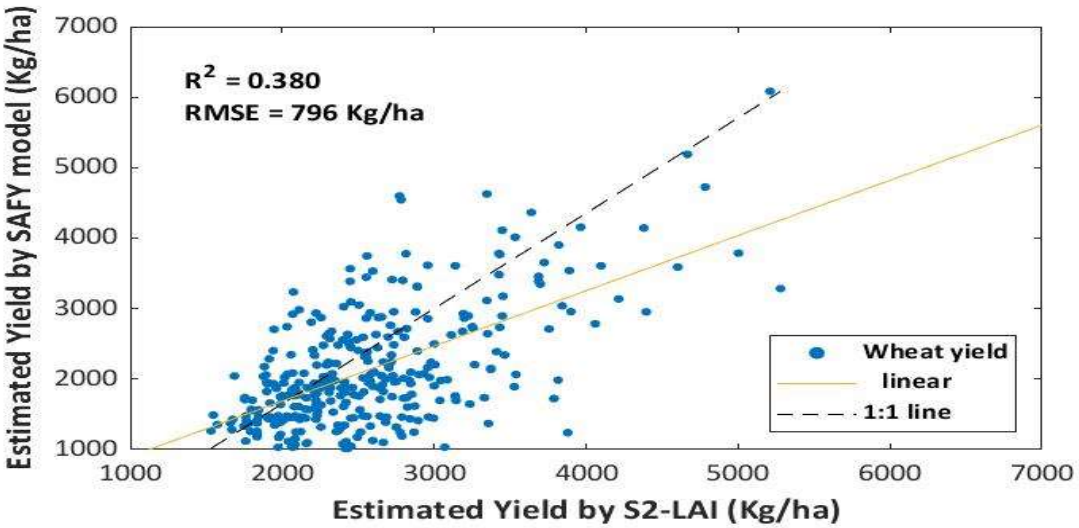


Figure 27: Comparison between yield estimated by S2-LAI and yield estimated by assimilation of LAI in the SAFY model.

Table 3 shows the summary of the SAFY model performance evaluation with and without assimilation. It includes the p-values obtained through the statistical significance testing for the three model simulation cases. The p-values are 0.00001, 0.00123, and 0.00001 respectively. Based on these results, the null hypothesis H_0 was rejected for all three model simulation cases and we conclude that the accuracy of yield estimation with the assimilation of RS data in the SAFY model significantly improves compared to the accuracy of yield estimation without the assimilation of RS data in the SAFY model.

Table 3: Summary of the SAFY model performance evaluation

		SAFY without assimilation		SAFY with assimilation		P-value
Cropfield	N	RMSE (Kg/ha)	R^2	RMSE (Kg/ha)	R^2	
Rainfed	557	1325	0.003	744	0.292	0.00001
Irrigated	86	3289	0.040	1786	0.152	0.00123
Combination of rainfed and irrigated	643	1722	0.008	952	0.378	0.00001

5.3 Discussion and comparison with prior research

The initial phase of this thesis involved incorporating meteorological time-series data into the SAFY model to estimate wheat yield and leaf area index (LAI). The findings presented in Figures 15 to 17 clearly demonstrate that, on the whole, the model tends to inaccurately represent the phenology of wheat LAI and underestimates wheat yield. When comparing the predicted and actual yield values, it became evident that the number of underestimated yield values is more than the overestimated ones, especially for yield values ranging from 2000 to 7000 kg/ha. In this range, the error exceeded the average RMSE. The substantial discrepancy arises from the SAFY model's inability to accurately capture the dynamics of wheat phenology as reflected in the LAI compared to real-world observations.

It is important to mention that the estimation of the biophysical variable, LAI, in this study relied on a neural network algorithm. However, due to the unavailability of ground measurements, it was not possible to verify the accuracy of the LAI estimates. Consequently, since the LAI estimation was intended for assimilation purposes in yield estimation, it is plausible that errors in the LAI estimates could have influenced the results of the assimilation process. However, the smoothness of the retrieved LAI along the season illustrates the low

noise level produced by the BV-NET algorithm. Furthermore, Delloye et al. (2019) have conducted a precise validation of the retrieved LAI, specifically for wheat. These results imply that field-based LAI measurements may exhibit a comparable level of error to the remotely sensed LAI.

Previous studies, such as those conducted by Ma et al. (2022) and Han et al. (2021), have reported that incorporating remote sensing LAI into crop yield models through assimilation methods leads to improved estimates of wheat yield. Ma et al. (2022) assert that the SAFY model has a significant advantage over more complex models like WOFOST and AquaCrop. It offers a more straightforward method for crop growth modelling, effectively addressing the drawbacks of other models' complex parameterization. SAFY manages to retain the main benefits of crop growth models while making the process easier to use and more adaptable to a variety of situations.

In this research, we employed daily time-series RS-LAI data derived from S2 imagery for each crop field and utilized the DE optimization algorithm to optimize the free parameters in the SAFY model. The results, as depicted in Figure 18, illustrate an excellent agreement between the time-series of SAFY-simulated LAI and the S2-LAI data. Figures 19 and 20 present a comparison of the results for rainfed, irrigated, and combined crop fields. It is observed that a good relationship exists between the two data in all scenarios. Moreover, for the combined rainfed and irrigated crop fields, the average R^2 and RMSE of the LAI were found to be 0.967 and 0.201, respectively. These findings indicate that the SAFY model is capable of accurately simulating the seasonal pattern of LAI, thus validating the optimization method and the accuracy of the phenological parameters which were calibrated in this study.

The accuracy of the LAI simulations obtained in this study also aligns with a previous investigation conducted by Dong et al. (2016), which reported an R^2 of 0.96 and an RMSE of 0.23 when assimilating LAI derived from remote sensing data into the SAFY model.

Figure 21 and 22 depicts the comparison between the actual and estimated yield. The results indicate that the assimilation of the S2-LAI in the model slightly improved the wheat yield estimation with respect to the R^2 and RMSE values compared to without assimilation. The actual wheat yield data ranged from 1100 to 6500 kg/ha compared to the SAFY simulated yield that ranged from 1016 to 6083 kg/ha. However, the variation seen in the actual wheat yield data was not sufficiently accounted for by the SAFY model as shown in Figure 22. As a result, the model tends to overestimate the low yield and underestimate the high yield value. A similar

situation was also reported by Han et al. (2021). The overall estimated wheat yield in this study shows a moderate agreement with the actual wheat yield data with an R^2 value of 0.38 and RMSE of 952 kg/ha. This suggests that the SAFY crop growth model, when coupled with remote sensing biophysical variables, has the potential to be used for predicting wheat yield. Despite the fact that the model could not account for the variation within the cropfield in this study, the correlation agreement and estimation error showed improved performance in comparison to earlier studies that made use of the SAFY model to predict wheat yield as shown in Table 4.

The results shown in Figure 27 demonstrate a moderate agreement between the yield estimates obtained by the linear regression model and the SAFY model. The average estimated yield obtained from the regression model is 2530 Kg/ha, which is close to the average observed yield of 2571 Kg/ha, while the SAFY model produces an average estimated yield of 2122 Kg/ha. Furthermore, the regression model could also not effectively account for the variability in the field-measured yield with the estimated yield in the range of 1499 to 5455 kg/ha. Thus, the model also tends to overestimate the low wheat yield values and underestimate the high yield values as seen in Figure 26.

The obtained P-values from the F-statistics also provide evidence of the statistical significance of the SAFY model's performance when assimilating remote sensing data. These results further validate the model's capability to effectively improve yield estimation in the study area.

Table 4: Performance in terms of correlation and estimation error between this study and earlier studies that focused on wheat yield estimation using SAFY model

Reference	EO source	R^2	RMSE (kg/ha)	Yield range (kg/ha)
Han et al. (2021)	MODIS	0.30		2500 - 10500
Manivasagam et al. (2021)	S2	0.35	880	3110 - 8010
Gaso et al. (2019)	Landsat-7 ETM and Landsat-8 OLI	0.11	1532	1600 - 5600
Silvestro et al. (2017)	HJ1A/B, Landsat-8 OLI		1090	2500 - 8500
Dong et al. (2016) (biomass)	Landsat-8 OLI and MODIS		1760 - 2310	3000 - 16000

The findings of this study highlight that S2 satellite data offers high spatial and temporal information that can be utilized to derive crop variables, such as LAI. Incorporating this derived LAI data into the SAFY model has shown potential for enhancing the estimation of wheat yield. However, there are likely four reasons that contributed to the low R^2 value obtained in this study and also the underestimation and overestimation of some yield values: (1) The day of plant emergence (D_0) used in this research is not exactly the one observed in the field and can affect the model estimation as agricultural conditions vary field by field. It would have been helpful to derive this information from the EO for each cropfield to simulate the crop yield because the growth of the late crop and the proportion of early emergence can both be significantly impacted by choosing a day that is either a bit early or late for the crop (Ma et al., 2022).

(2) The model does not incorporate all the impacts of environmental stresses on crop yield (except temperature) and management practices and this may result in the poor performance of the model (Manivasagam et al., 2021). It would have also been helpful to have the management information for the cropfields as this could provide further insights into crop performance and yield estimation, however, this study was limited by the unavailability of these data for inclusive analysis.

(3) The yield data utilized in this study were pre-processed to enhance their quality. It is important to note that these in-situ measurements were not considered perfect due to the methods used in estimating the yield, which are prone to errors. These methods include eye estimation, counting and weighing of spikes, and counting of spikes and grains. Additionally, only few of the cropfields yield information were provided directly by the farmers. To mitigate the potential influence of heterogeneous field conditions, yield data with significant variability were identified and excluded based on the maximum S2-LAI. However, it is important to note that all errors were not removed by the outlier test that was conducted, which may have resulted in bias in the statistical yield results, particularly in relation to the yield validation data. Specifically, the presence of this residual bias increased the variation between the actual and estimated yield. Additionally, the limitations of S2 image pixels in capturing small field features may have also contributed to this variation since farmer information on crop cutting of $1m^2$ is not necessarily representative of a whole field.

(4) This study used the differential evolution algorithm to assimilate the LAI and optimize the free parameters in the SAFY model by minimizing the error between the S2-LAI and SAFY

simulated LAI. Future research could investigate the assimilation of LAI and utilize yield data to optimize the SAFY model parameters using popular optimization algorithms like particle swarm optimization (PSO), Markov-chain Monte Carlo (MCMC) optimization, and data assimilation method by forcing and updating method which will require having field measured data.

6. CONCLUSION

This master's thesis focused on improving wheat yield estimation by assimilating remote sensing biophysical variables, specifically LAI derived from S2 satellite images for the year 2019, into the SAFY model. Prior to the assimilation of LAI, meteorological time-series data was ingested into the model to estimate wheat yield for rainfed, irrigated, and a combination of both cropfields. However, the results showed that the model was ineffective to depict the wheat phenology. The comparison between the actual and simulated yield values shows no correlation for the three model simulation cases, primarily due to the misrepresentation of the wheat phenology and non-representativeness of the default parameters with respect to the simulation. To address this limitation, assimilation of LAI retrieved from S2 images was performed, utilizing the differential evolution algorithm to optimize six free parameters and minimize the error between the remotely sensed LAI and the SAFY simulated LAI.

Following the assimilation process, the results showed significant improvements for the three model simulation cases. The SAFY model, when integrated with the remote sensing LAI, demonstrated the capability to accurately simulate the dynamic range of LAI and capture the seasonal phenology of the wheat crop. Then, a linear relationship between the remote sensing LAI and field-measured yield is established. The result indicates a moderate correlation between the two variables. Validation of the analysis using the LAI simulated by the SAFY model reveals that this method is also suitable for estimating crop yield.

Although some limitations remained, such as the model's inability to capture the variation in the actual wheat yield data and account for environmental stress effects, the SAFY model showed good performance and consistency with previous research that adopt the model in estimating wheat yield. It should be noted that further investigation is required to examine the impact of irrigation situations on the grain filling factor as this was not considered in this research and other factors such as soil characteristics and management practices which influence yield production should be considered for future work. In spite of this, utilising the remote sensing technology with crop growth models was proved efficient by this research to improve the estimation of wheat yield at field level.

Bibliography

Ansarifar, J., Wang, L., & Archontoulis, S. V. (2021). An interaction regression model for crop yield prediction. *Scientific Reports*, 11(1). http://dx.doi.org/10.1038/s41598-021-97221-7

Atzberger, C., & Richter, K. (2012). Spatially constrained inversion of radiative transfer models for improved LAI mapping from future Sentinel-2 imagery. *Remote Sensing of Environment*, 120, 208–218. http://dx.doi.org/10.1016/j.rse.2011.10.035

Atzberger, C., Darvishzadeh, R., Immitzer, M., Schlerf, M., Skidmore, A., & le Maire, G. (2015). Comparative analysis of different retrieval methods for mapping grassland leaf area index using airborne imaging spectroscopy. *International Journal of Applied Earth Observation and Geoinformation*, 43, 19–31. http://dx.doi.org/10.1016/j.jag.2015.01.009

Awika, J. M. (2011). Major Cereal Grains Production and Use around the World. Advances in Cereal Science: *Implications to Food Processing and Health Promotion*, 1–13. https://doi.org/10.1021/bk-2011-1089.ch001

Bai, T., Zhang, N., Mercatoris, B., & Chen, Y. (2019). Improving Jujube Fruit Tree Yield Estimation at the Field Scale by Assimilating a Single Landsat Remotely Sensed LAI into the WOFOST Model. *Remote Sensing*, 11(9), 1119. https://doi.org/10.3390/rs11091119

Baidar, T. (2020). Rice Crop Classification and Yield Estimation Using Multi-Temporal Sentinel-2 Data: A Case Study of Terai Districts of Nepal. Universitat Jaume I (UJI), *MSc. GeoTech Thesis*

Balaghi, R., Tychon, B., Eerens, H., & Jlibene, M. (2008). Empirical regression models using NDVI, rainfall and temperature data for the early prediction of wheat grain yields in Morocco. *International Journal of Applied Earth Observation and Geoinformation*, 10(4), 438–452. http://dx.doi.org/10.1016/j.jag.2006.12.001

Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., & Sinoquet, H. (2003). An overview of the crop model stics. *European Journal of Agronomy*, 18(3-4), 309–332. doi:10.1016/s1161-0301(02)00110-7

Casa, R., Rossi, M., & Sappa, G. (2009). Assessing Crop Water Demand by Remote Sensing and GIS for the Pontina Plain, Central Italy. *Water Resource Management*, 23, 1685–1712. https://doi.org/10.1007/s11269-008-9347-4

Casa, R., Varella, H., Buis, S., Guérif, M., De Solan, B., & Baret, F. (2012). Forcing a wheat crop model with LAI data to access agronomic variables: Evaluation of the impact of model and LAI uncertainties and comparison with an empirical approach. *European Journal of Agronomy*, 37(1), 1–10. https://doi.org/10.1016/j.eja.2011.09.004

Chahbi, A., Zribi, M., Lili-Chabaane, Z., Duchemin, B., Shabou, M., Mougenot, B., & Boulet, G. (2014). Estimation of the dynamics and yields of cereals in a semi-arid area using remote sensing and the SAFY growth model. *International Journal of Remote Sensing*, 35(3), 1004–1028. http://dx.doi.org/10.1080/01431161.2013.875629

Chen, J.M., Menges, C.H., & Leblanc, S.G. (2005). Global mapping of foliage clumping index using multi-angular satellite data. *Remote Sensing of Environment*, 97, 447-457

Claverie, M., Demarez, V., Duchemin, B., Hagolle, O., Ducrot, D., Marais-Sicre, C., & Dedieu, G. (2012). Maize and sunflower biomass estimation in southwest France using high spatial and

temporal resolution remote sensing data. *Remote Sensing of Environment*, 124, 844–857. http://dx.doi.org/10.1016/j.rse.2012.04.005

Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., & Wang, L. (2003). Retrieval of canopy biophysical variables from bidirectional reflectance. Remote Sensing of Environment, 84(1), 1–15. http://dx.doi.org/10.1016/s0034-4257(02)00035-4

Cui, Z., & Kerekes, J. P. (2018). Potential of Red Edge Spectral Bands in Future Landsat Satellites on Agroecosystem Canopy Green Leaf Area Index Retrieval. *Remote Sensing*, 10(9), 1458. http://dx.doi.org/10.3390/rs10091458

Curnel, Y., de Wit, A. J. W., Duveiller, G., & Defourny, P. (2011). Potential performances of remotely sensed LAI assimilation in WOFOST model based on an OSS Experiment. *Agricultural and Forest Meteorology,* 151(12), 1843–1855. 10.1016/j.agrformet.2011.08.002

Dahms, T., Seissiger, S., Conrad, C., & Borg, E. (2016). Modelling biophysical parameters of maize using landsat 8 time series. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B2, 171–175, https://doi.org/10.5194/isprs-archives-XLI-B2-171-2016

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 Sensing of Environment*, 216, 245–261. doi:10.1016/j.rse.2018.06.037

De Wit, A. J. W., & van Diepen, C. A. (2007). Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. *Agricultural and Forest Meteorology*, 146(1-2), 38–56. https://doi.org/10.1016/j.agrformet.2007.05.004

De Wit, A., Duveiller, G., & Defourny, P. (2012). Estimating regional winter wheat yield with WOFOST through the assimilation of green area index retrieved from MODIS observations. *Agricultural and Forest Meteorology*, 164, 39–52. doi:10.1016/j.agrformet.2012.04.011

Delécolle, R., Maas, S. J., Guérif, M., & Baret, F. (1992). Remote sensing and crop production models: present trends. *ISPRS Journal of Photogrammetry and Remote Sensing*, 47(2-3), 145–161. http://dx.doi.org/10.1016/0924-2716(92)90030-d

Diepen, C. A., Wolf, J., Keulen, H., & Rappoldt, C. (1989). WOFOST: a simulation model of crop production. *Soil Use and Management*, 5(1), 16–24. http://dx.doi.org/10.1111/j.1475-2743.1989.tb00755.x

Dong, T., Liu, J., Qian, B., Zhao, T., Jing, Q., Geng, X., & Shang, J. (2016). Estimating winter wheat biomass by assimilating leaf area index derived from fusion of Landsat-8 and MODIS data. *International Journal of Applied Earth Observation and Geoinformation*, 49, 63–74. http://dx.doi.org/10.1016/j.jag.2016.02.001

Donohue, R. J., Lawes, R. A., Mata, G., Gobbett, D., & Ouzman, J. (2018). Towards a national, remote-sensing-based model for predicting field-scale crop yield. *Field Crops Research*, 227, 79–90. https://doi.org/10.1016/j.fcr.2018.08.005

Dorigo, W. A., Zurita-Milla, R., de Wit, A. J. W., Brazile, J., Singh, R., & Schaepman, M. E. (2007). A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation,* 9(2), 165–193. https://doi.org/10.1016/j.jag.2006.05.003

Dorigo, W., Richter, R., Baret, F., Bamler, R., & Wagner, W. (2009). Enhanced Automated Canopy Characterization from Hyperspectral Data by a Novel Two Step Radiative Transfer Model Inversion Approach. *Remote Sensing*, 1(4), 1139–1170. doi:10.3390/rs1041139

Duchemin, B., Maisongrande, P., Boulet, G., & Benhadj, I. (2008). A simple algorithm for yield estimates: Evaluation for semi-arid irrigated winter wheat monitored with green leaf area index. *Environmental Modelling & Software*, 23(7), 876–892. 10.1016/j.envsoft.2007.10.003

Duveiller, G., Weiss, M., Baret, F., & Defourny, P. (2011). Retrieving wheat Green Area Index during the growing season from optical time series measurements based on neural network radiative transfer inversion. *Remote Sensing of Environment*, 115(3), 887–896.

Dzotsi, K. A., Basso, B., & Jones, J. W. (2015). Parameter and uncertainty estimation for maize, peanut and cotton using the SALUS crop model. *Agricultural Systems*, 135, 31–47. http://dx.doi.org/10.1016/j.agsy.2014.12.003

ESA (2015). Spatial Resolution. https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial

ESA Sen2Agri (2020). Software User Manual. Software User Manual (esa-sen2agri.org)

Esri, HERE, Garmin, FAO, NOAA, & USGS (2012). World terrain basemap.

Fang, H., Liang, S., & Hoogenboom, G. (2011). Integration of MODIS LAI and vegetation index products with the CSM–CERES–Maize model for corn yield estimation. International *Journal of Remote Sensing*, 32(4), 1039–1065. http://dx.doi.org/10.1080/01431160903505310

FAO (2000). The State of Food and Agriculture, 2000. www.fao.org

Gaso, D. V., Berger, A. G., & Ciganda, V. S. (2019). Predicting wheat grain yield and spatial variability at field scale using a simple regression or a crop model in conjunction with Landsat images. *Computers and Electronics in Agriculture*, 159, 75–83.

Georgioudakis, M., & Plevris, V. (2020). A Comparative Study of Differential Evolution Variants in Constrained Structural Optimization. *Frontiers in Built Environment*, 6. doi:10.3389/fbuil.2020.00102

Gilardelli, C., Stella, T., Confalonieri, R., Ranghetti, L., Campos-Taberner, M., García-Haro, F. J., & Boschetti, M. (2019). Downscaling rice yield simulation at sub-field scale using remotely sensed LAI data. *European Journal of Agronomy*, 103, 108–116.

Gozdowski, D., Samborski, S., & Dobers, E. (2010). Evaluation of methods for the detection of spatial outliers in the yield data of winter wheat. *Colloquium Biometricum*, 40, 41-51

Gu, Y., Bélair, S., Mahfouf, J.F., & Deblonde, G. (2006). Optimal interpolation analysis of leaf area index using MODIS data. *Remote Sensing of Environment*, 104(3), 283–296. doi:10.1016/j.rse.2006.04.021

Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., & Strachan, I.B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, 90(3), 337–352. http://dx.doi.org/10.1016/j.rse.2003.12.013

Han, D., Wang, P., Tansey, K., Zhang, S., Tian, H., & Zhang, Y. (2021). Improving wheat yield estimates by integrating a remotely sensed drought monitoring index into the Simple Algorithm for Yield Estimate Model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 10383-10394, doi: 10.1109/JSTARS.2021.3119398.

Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., & Keating, B. A. (2014). APSIM – Evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software*, 62, 327–350.

https://cds.climate.copernicus.eu/

https://ipad.fas.usda.gov/rssiws/al/crop_calendar/europe.aspx

https://www.mapa.gob.es/

https://www.statology.org/f-distribution-calculator

Hu, S., Shi, L., Huang, K., Zha, Y., Hu, X., Ye, H., & Yang, Q. (2019). Improvement of sugarcane crop simulation by SWAP-WOFOST model via data assimilation. *Field Crops Research*, 232, 49–61. https://doi.org/10.1016/j.fcr.2018.12.009

Huang, J., Ma, H., Sedano, F., Lewis, P., Liang, S., Wu, Q., & Zhu, D. (2019). Evaluation of regional estimates of winter wheat yield by assimilating three remotely sensed reflectance datasets into the coupled WOFOST–PROSAIL model. *European Journal of Agronomy*, 102, 1–13. https://doi.org/10.1016/j.eja.2018.10.008

Huang, J., Ma, H., Su, W., Zhang, X., Huang, Y., Fan, J., & Wu, W. (2015). Jointly Assimilating MODIS LAI and ET Products into the SWAP Model for Winter Wheat Yield Estimation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(8), 4060–4071. http://dx.doi.org/10.1109/jstars.2015.2403135

Jiang, Z., Chen, Z., Chen, J., Ren, J., Li, Z., & Sun, L. (2014). The Estimation of Regional Crop Yield Using Ensemble-Based Four-Dimensional Variational Data Assimilation. *Remote Sensing*, 6(4), 2664–2681. http://dx.doi.org/10.3390/rs6042664

Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., & Wang, J. (2018). A review of data assimilation of remote sensing and crop models. *European Journal of Agronomy*, 92, 141–152. http://dx.doi.org/10.1016/j.eja.2017.11.002

Jin, X., Kumar, L., Li, Z., Xu, X., Yang, G., & Wang, J. (2016). Estimation of Winter Wheat Biomass and Yield by Combining the AquaCrop Model and Field Hyperspectral Data. *Remote Sensing*, 8(12), 972. https://doi.org/10.3390/rs8120972

Jones, J., Hoogenboom, G., Porter, C., Boote, K., Batchelor, W., Hunt, L., & Ritchie, J. (2003). The DSSAT cropping system model. *European Journal of Agronomy*, 18(3-4), 235–265. https://doi.org/10.1016/s1161-0301(02)00107-7

Kalpana, R., Natarajan, S., Mythili, S.R., Shekinah, D.E., & Krishnarajan, J. (2003). Remote sensing for crop monitoring – A review. *Agricultural Reviews, 24*, 31-39.

Khaki, S., & Wang, L. (2019). Crop Yield Prediction Using Deep Neural Networks. *Frontiers in Plant Science*, 10. http://dx.doi.org/10.3389/fpls.2019.00621

Kim, N., & Lee, Y. (2014). Estimation of corn and soybeans yield using remote sensing and crop yield data in the United States. *Remote Sensing for Agriculture, Ecosystems, and Hydrology* XVI. http://dx.doi.org/10.1117/12.2067311

Kimes, D. S., Knyazikhin, Y., Privette, J. L., Abuelgasim, A. A., & Gao, F. (2000). Inversion methods for physically-based models. *Remote Sensing Reviews*, 18(2-4), 381–439. http://dx.doi.org/10.1080/02757250009532396

Li, F., Ren, J., Wu, S., Zhao, H., & Zhang, N. (2021). Comparison of Regional Winter Wheat Mapping Results from Different Similarity Measurement Indicators of NDVI Time Series and Their Optimized Thresholds. *Remote Sensing*, 13(6), 1162. https://doi.org/10.3390/rs13061162

Li, H., Chen, Z., Liu, G., Jiang, Z., & Huang, C. (2017). Improving Winter Wheat Yield Estimation from the CERES-Wheat Model to Assimilate Leaf Area Index with Different Assimilation Methods and Spatio-Temporal Scales. *Remote Sensing*, 9(3), 190. https://doi.org/10.3390/rs9030190

Li, Z., Jin, X., Zhao, C., Wang, J., Xu, X., Yang, G., & Shen, J. (2015). Estimating wheat yield and quality by coupling the DSSAT-CERES model and proximal remote sensing. *European Journal of Agronomy*, 71, 53–62. http://dx.doi.org/10.1016/j.eja.2015.08.006

Li, X., Zhang, Y., Bao, Y., Luo, J., Jin, X., Xu, X., & Yang, G. (2014). Exploring the best hyperspectral features for LAI estimation using partial least squares regression. *Remote Sensing*, 6(7), 6221–6241. doi:10.3390/rs6076221

Lillesand, T, M. Kiefer, R.W. & Chipman, J.W. (2004). Remote sensing and Image interpretation. Fifth edition. John and Sons, Inc. New York. 828 p.

Lobell, D. B. (2013). The use of satellite data for crop yield gap analysis. *Field Crops Research*, 143, 56–64. https://doi.org/10.1016/j.fcr.2012.08.008

Ma, C., Liu, M., & Ding, F. (2022). Wheat growth monitoring and yield estimation based on remote sensing data assimilation into the SAFY crop growth model. *Scientific Reports*, 12, 5473. https://doi.org/10.1038/s41598-022-09535-9

Maas, S. J. (1993). Parameterized Model of Gramineous Crop Growth: Leaf Area and Dry Mass Simulation. *Agronomy Journal*, 85(2), 348. doi:10.2134/agronj1993.00021962008500002

Mananze, S., Pôças, I., & Cunha, M. (2018). Retrieval of Maize Leaf Area Index Using Hyperspectral and Multispectral Data. *Remote Sensing*, 10(12), 1942.

Manivasagam, V. S., & Rozenstein, O. (2020). Practices for upscaling crop simulation models from field scale to large regions. *Computers and Electronics in Agriculture,* 175, 105554. https://doi.org/10.1016/j.compag.2020.105554

Manivasagam, V. S., Sadeh, Y., Kaplan, G., Bonfil, D. J., & Rozenstein, O. (2021). Studying the Feasibility of Assimilating Sentinel-2 and PlanetScope Imagery into the SAFY Crop Model to Predict Within-Field Wheat Yield. *Remote Sensing*, 13(12), 2395.

Monteith, J. L. (1972). Solar Radiation and Productivity in Tropical Ecosystems. *The Journal of Applied Ecology*, 9(3), 747. http://dx.doi.org/10.2307/2401901

Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114(4), 358–371.

Nouvellon, Y., Moran, M. S., Seen, D. L., Bryant, R., Rambal, S., Ni, W., & Qi, J. (2001). Coupling a grassland ecosystem model with Landsat imagery for a 10-year simulation of carbon and water budgets. *Remote Sensing of Environment*, 78(1-2), 131–149. http://dx.doi.org/10.1016/s0034-4257(01)00255-3

Novelli, F., & Vuolo, F. (2019). Assimilation of Sentinel-2 Leaf Area Index Data into a Physically Based Crop Growth Model for Yield Estimation. *Agronomy*, 9(5), 255. http://dx.doi.org/10.3390/agronomy9050255

Padilla, F. L. M., Maas, S. J., González-Dugo, M. P., Mansilla, F., Rajan, N., Gavilán, P., & Domínguez, J. (2012). Monitoring regional wheat yield in Southern Spain using the GRAMI model and satellite imagery. *Field Crops Research*, 130, 145–154.

Pan, H., & Chen, Z. (2021). Crop Growth Modeling and Yield Forecasting. In: L. Di, B. Üstündağ (eds.), *Agro-geoinformatics,* Springer Remote Sensing/Photogrammetry, https://doi.org/10.1007/978-3-030-66387-2_11

Pan, H., Chen, Z., Ren, J., Li, H., & Wu, S. (2019). Modeling winter wheat leaf area index and canopy water content with three different approaches using Sentinel-2 multispectral instrument data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(2), 482–492, doi:10.1109/JSTARS.2018.2855564.

Pellenq, J. & Boulet, G. (2004). A methodology to test the pertinence of remote-sensing data assimilation into vegetation models for water and energy exchange at the land surface. http://dx.doi.org/10.1051/agro:2004017

Peng, X., Han, W., Ao, J., & Wang, Y. (2021). Assimilation of LAI Derived from UAV Multispectral Data into the SAFY Model to Estimate Maize Yield. *Remote Sensing*, 13(6), 1094. https://doi.org/10.3390/rs13061094

Pôças, I., Gonçalves, J., Costa, P. M., Gonçalves, I., Pereira, L. S., & Cunha, M. (2017). Hyperspectral-based predictive modelling of grapevine water status in the Portuguese Douro wine region. *International Journal of Applied Earth Observation and Geoinformation*, 58, 177–190. http://dx.doi.org/10.1016/j.jag.2017.02.013

Population Reference Bureau (2020). 2020 World Population Data Sheet

Portillo, J.E. (2016). Assimilation of Leaf Area Index into the crop growth models GRAMI and SAFY to monitor maize crop: a case of study in the Marchfeld region. University of Natural Resources and Life Sciences, Vienna (BOKU), *PhD thesis*.

Raes, D., Steduto, P., Hsiao, T. C., & Fereres, E. (2009). AquaCrop: The FAO Crop Model to Simulate Yield Response to Water: II. Main Algorithms and Software Description. *Agronomy Journal*, 101(3), 438. doi:10.2134/agronj2008.0140s

Rembold, F., Atzberger, C., Savin, I., & Rojas, O. (2013). Using low resolution satellite imagery for yield prediction and yield anomaly detection. *Remote Sensing*, 5(11), 5572–5573. http://dx.doi.org/10.3390/rs5115572

Ren, J., Chen, Z., Zhou, Q., & Tang, H. (2008). Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China. *International Journal of Applied Earth Observation and Geoinformation*, 10(4), 403–413. http://dx.doi.org/10.1016/j.jag.2007.11.003

Ritchie, J. T. (1985). A user-orientated model of the soil water balance in wheat. Wheat growth and modelling, 293-305.

Rivera, J., Verrelst, J., Delegido, J., Veroustraete, F., & Moreno, J. (2014). On the semi-automatic retrieval of biophysical parameters based on spectral index optimization. *Remote Sensing*, 6(6), 4927–4951. http://dx.doi.org/10.3390/rs6064927

Rivera, J., Verrelst, J., Leonenko, G., & Moreno, J. (2013). Multiple cost functions and regularization options for improved retrieval of leaf chlorophyll content and LAI through inversion of the PROSAIL model. *Remote Sensing*. 5, 3280–3304.

Rouse, J.W., Haas, R.H., Schell, J.A. & Deering, D.W. (1973). Monitoring vegetation systems in the Great Plains with ERTS. *Proceedings of the Third ERTS Symposium*, Washington DC, 10-14 December 1973, 309-317.

Royo, C. & Briceño-Felix, G. (2011). Wheat breeding in Spain. *The World Wheat Book*: *A History of Wheat Breeding*. 2, 121-154.

Sabater, J. (2019). ERA5-Land monthly averaged data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). 10.24381/cds.68d2bb30

Sadegh, M., & Vrugt, J.A. (2014). Approximate Bayesian computation using Markov Chain Monte Carlo simulation: DREAM(ABC). *Water Resources Research*, 50, 6767 - 6787. https://doi.org/10.1002/2014WR015386

Segl, K., Richter, R., Küster, T., & Kaufmann, H. (2012). End-to-end sensor simulation for spectral band selection and optimization with application to the Sentinel-2 mission. *Applied Optics*, 51(4), 439. doi:10.1364/ao.51.000439

Shiferaw, B., Smale, M., Braun, H.J., Duveiller, E., Reynolds, M., & Muricho, G. (2013). Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Security*, 5(3), 291–317. 10.1007/s12571-013-0263-y

Silvestro, P., Pignatti, S., Pascucci, S., Yang, H., Li, Z., Yang, G., & Casa, R. (2017). Estimating wheat yield in China at the field and district Scale from the assimilation of satellite data into the Aquacrop and Simple Algorithm for Yield (SAFY) Models. *Remote Sensing*, 9(5), 509. http://dx.doi.org/10.3390/rs9050509

Son, Nguyen-Thanh & Chen, C. & Chen, C. & Chang, L. & Chiang, S. (2016). Rice yield estimation through assimilating satellite data into a crop simulation model. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. XLI-B8. 993-996. http://dx.doi.org/10.5194/isprs-archives-XLI-B8-993-2016

Steduto, P., Hsiao, T. C., Raes, D., & Fereres, E. (2009). AquaCrop—The FAO Crop Model to Simulate Yield Response to Water: I. Concepts and Underlying Principles. *Agronomy Journal*, 101(3), 426. http://dx.doi.org/10.2134/agronj2008.0139s

Stöckle, C. O., Donatelli, M., & Nelson, R. (2003). CropSyst, a cropping systems simulation model. *European Journal of Agronomy*, 18(3-4), 289–307. 10.1016/s1161-0301(02)00109-0

Storn, R., & Price, K. (1997). Differential Evolution – A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4), 341–359. doi:10.1023/a:1008202821328

Thorp, K.R., Hunsaker, D.J., & French, A.N. (2010). Assimilating leaf area index estimates from remote sensing into the simulations of a cropping systems model. *Transactions of the ASABE*, 53, 251-262.

Tripathy, R., Chaudhari, K. N., Mukherjee, J., Ray, S. S., Patel, N. K., Panigrahy, S., & Parihar, J. S. (2013). Forecasting wheat yield in Punjab state of India by combining crop simulation model WOFOST and remotely sensed inputs. *Remote Sensing Letters*, 4(1), 19–28. https://doi.org/10.1080/2150704x.2012.683117

Tuvdendorj, B., Wu, B., Zeng, H., Batdelger, G., & Nanzad, L. (2019). Determination of appropriate remote sensing indices for spring wheat yield estimation in Mongolia. *Remote Sensing*, 11(21), 2568. https://doi.org/10.3390/rs11212568

United Nations (2015). "Transforming our world: The 2030 agenda for sustainable development". In: New York: United Nations, Department of Economic and Social Affairs.

Upreti, D., Pignatti, S., Pascucci, S., Tolomio, M., Huang, W., & Casa, R. (2020). Bayesian calibration of the Aquacrop-OS model for durum wheat by assimilation of canopy cover retrieved from VENuS satellite data. *Remote Sensing*, 12(16), 2666.

Vazifedoust, M., van Dam, J. C., Bastiaanssen, W. G. M., & Feddes, R. A. (2009). Assimilation of satellite data into agrohydrological models to improve crop yield forecasts. *International Journal of Remote Sensing*, 30(10), 2523–2545. doi:10.1080/01431160802552769

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J. P., Veroustraete, F., Clevers, J. G. P. W., & Moreno, J. (2015). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108, 273–290. http://dx.doi.org/10.1016/j.isprsjprs.2015.05.005

Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J. P., Camps-Valls, G., & Moreno, J. (2012). Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3. *Remote Sensing of Environment*, 118, 127–139. http://dx.doi.org/10.1016/j.rse.2011.11.002

Verrelst, J., Rivera, J. P., Veroustraete, F., Muñoz-Marí, J., Clevers, J. G. P. W., Camps-Valls, G., & Moreno, J. (2015). Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods – A comparison. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108, 260–272. http://dx.doi.org/10.1016/j.isprsjprs.2015.04.013

Vicente-Guijalba, F., Martinez-Marin, T., & Lopez-Sanchez, J. M. (2014). Crop phenology estimation using a multitemporal model and a Kalman filtering strategy. *IEEE Geoscience and Remote Sensing Letters*, 11(6), 1081–1085. https://doi.org/10.1109/lgrs.2013.2286214

Vuolo, F., Neuwirth, M., Immitzer, M., Atzberger, C., & Ng, W.T. (2018). How much does multi-temporal Sentinel-2 data improve crop type classification? *International Journal of Applied Earth Observation and Geoinformation,* 72, 122–130. doi:10.1016/j.jag.2018.06.007

Wang, H., Zhu, Y., Li, W., Cao, W., & Tian, Y. (2014). Integrating remotely sensed leaf area index and leaf nitrogen accumulation with RiceGrow model based on particle swarm optimization algorithm for rice grain yield assessment. *Journal of Applied Remote Sensing*, 8(1), 083674. http://dx.doi.org/10.1117/1.jrs.8.083674

Wang, X., Williams, J. R., Gassman, P. W., Baffaut, C., Izaurralde, R. C., Jeong, J., & Kiniry, J. R. (2012). EPIC and APEX: Model Use, Calibration, and Validation. *Transactions of the ASABE*, 55(4), 1447–1462. http://dx.doi.org/10.13031/2013.42253

Weiss, M., Baret, F., & Jay, S. (2020). S2ToolBox Level 2 products LAI, FAPAR, FCOVER. [*Research Report*] EMMAH-CAPTE, INRAe Avignon. ⟨hal-03584016⟩

Wu, Y., Xu, W., Huang, H., Huang, J., Yin, F., Ma, H., & Wang, X. (2020). Winter Wheat Yield Estimation at the Field Scale by Assimilating Sentinel-2 LAI into Crop Growth Model. IGARSS 2020 - 2020 *IEEE International Geoscience and Remote Sensing Symposium*. http://dx.doi.org/10.1109/igarss39084.2020.9323941

Xie, Y. & Wang, P. & Zhang, S. & Li, L. (2017). Winter Wheat Yield Estimation Based on Particle Filter Algorithm and Weights of Multi-variables. *Nongye Jixie Xuebao/Transactions of the Chinese Society for Agricultural Machinery.* 48, 148-155.

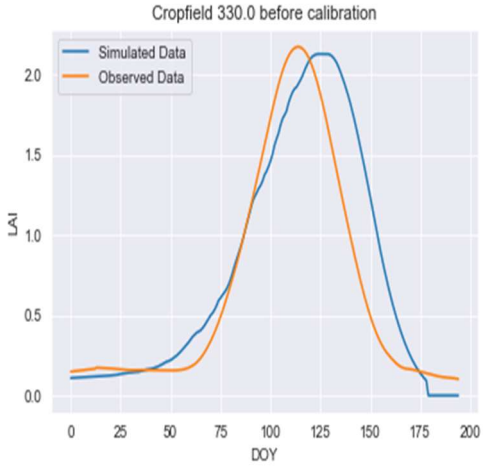
Xu, J.X., Ma, J., Tang, Y.N., Wu, W.X., Shao, J.H., Wu, W.B., & Guo, H.Q. (2020). Estimation of Sugarcane Yield Using a Machine Learning Approach Based on UAV-LiDAR Data. *Remote Sensing*, 12(17), 2823. https://doi.org/10.3390/rs12172823

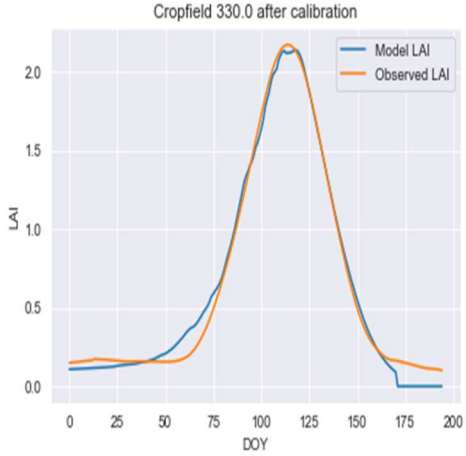
Yao, F., Tang, Y., Wang, P., & Zhang, J. (2015). Estimation of maize yield by using a process-based model and remote sensing data in the Northeast China Plain. *Physics and Chemistry of the Earth*, Parts A/B/C, 87-88, 142–152. https://doi.org/10.1016/j.pce.2015.08.010

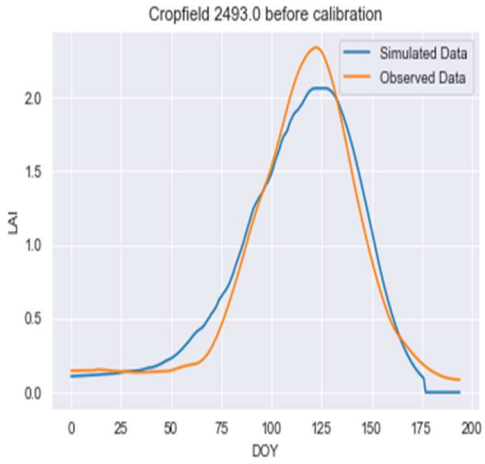
Zhao, Y., Han, S., Meng, Y., Feng, H., Li, Z., Chen, J., Song, X., Zhu, Y., & Yang, G. (2022). Transfer-Learning-Based Approach for Yield Prediction of Winter Wheat from Planet Data and SAFY Model. *Remote Sensing*, 14(21), 5474. https://doi.org/10.3390/rs14215474

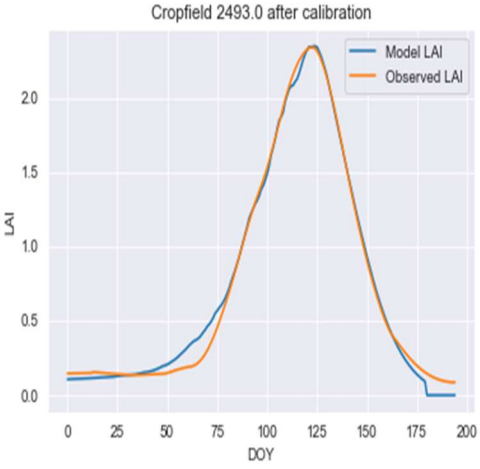
Zhou, X., Zheng, H. B., Xu, X. Q., He, J. Y., Ge, X. K., Yao, X., & Tian, Y. C. (2017). Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130, 246–255. http://dx.doi.org/10.1016/j.isprsjprs.2017.05.003

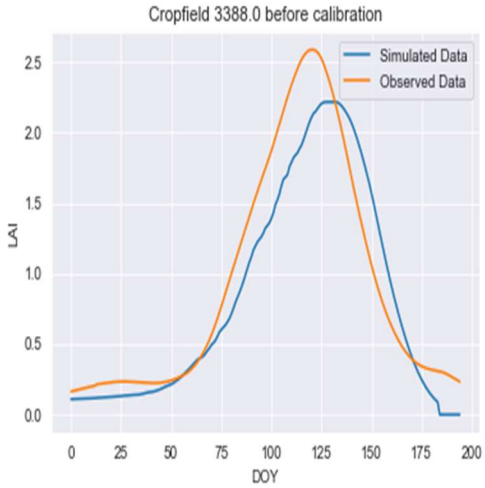
Appendix A: LAI simulation for selected cropfields using the best1bin DE method

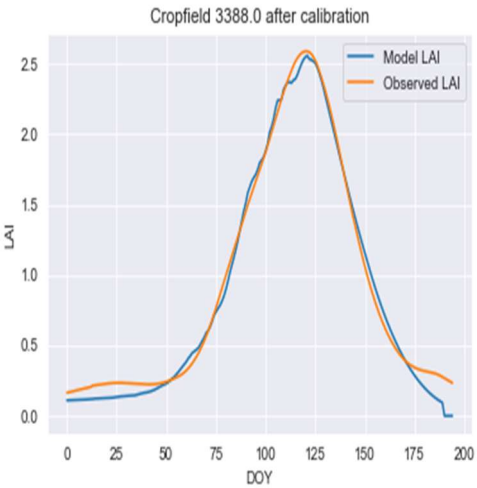


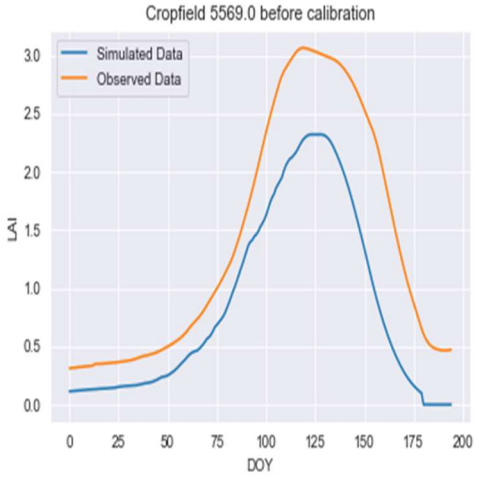


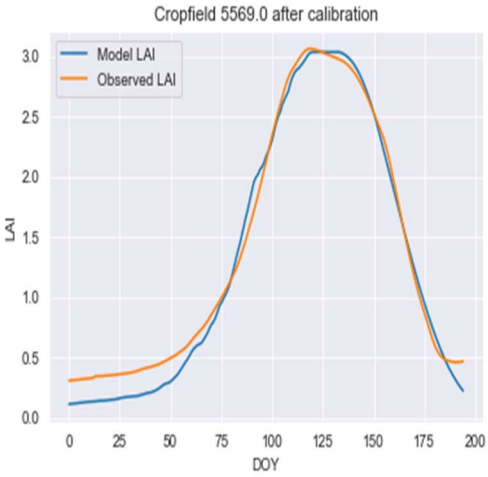


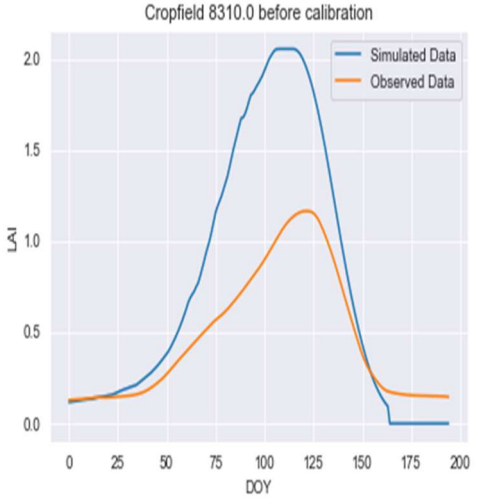


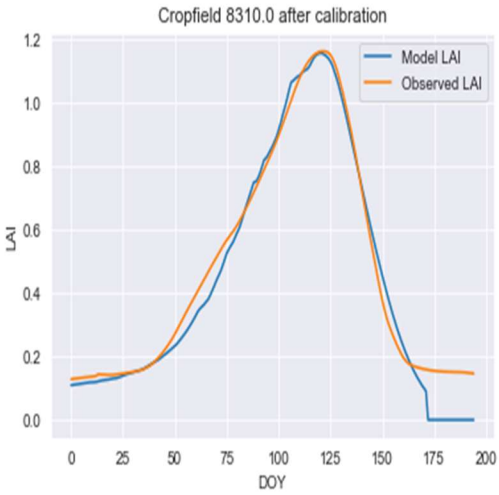


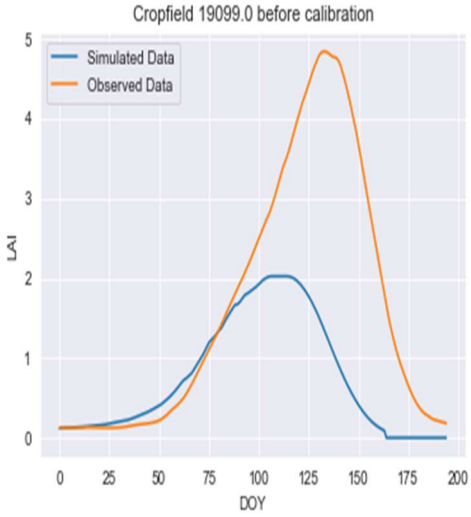


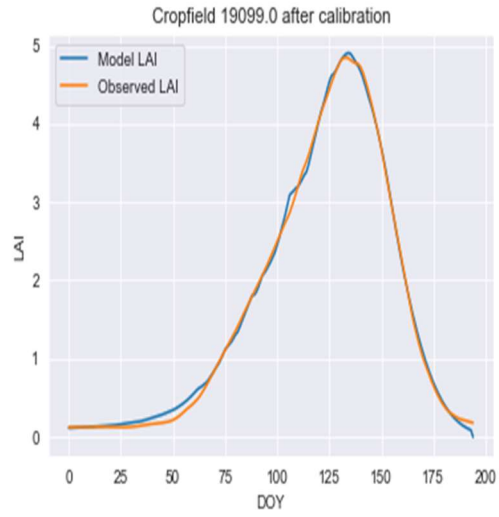

Cropfield 5569.0 before calibration
Simulated Data
Observed Data
LAI
DOY

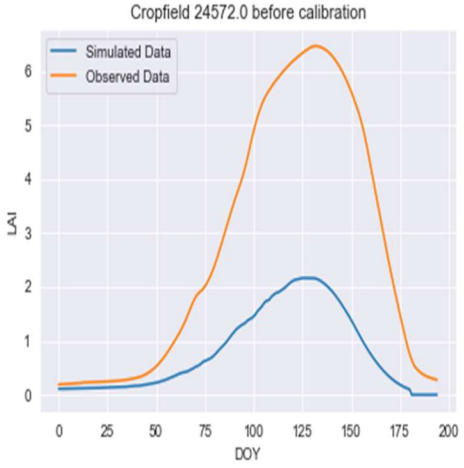

Cropfield 5569.0 after calibration
Model LAI
Observed LAI
LAI
DOY

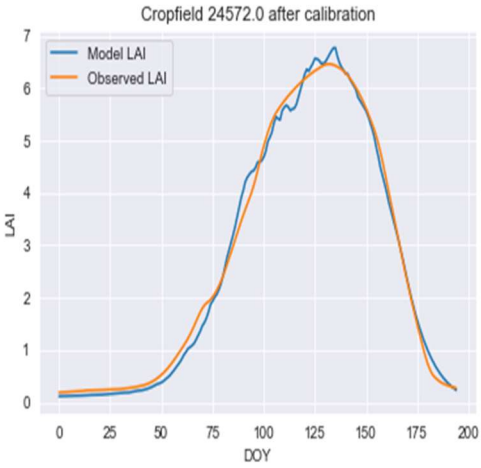

Cropfield 8310.0 before calibration
Simulated Data
Observed Data
LAI
DOY

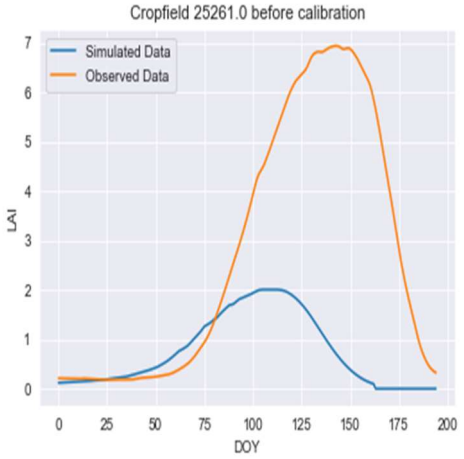

Cropfield 8310.0 after calibration
Model LAI
Observed LAI
LAI
DOY

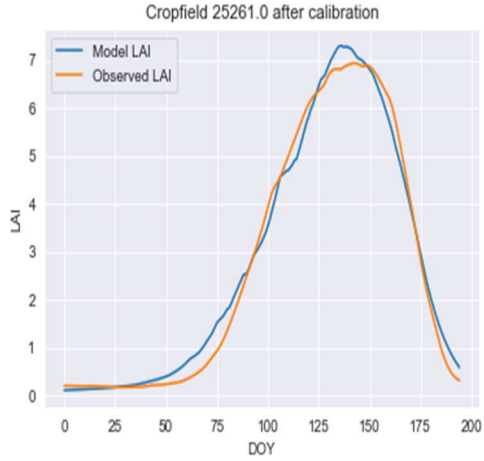

Cropfield 19099.0 before calibration
Simulated Data
Observed Data
LAI
DOY

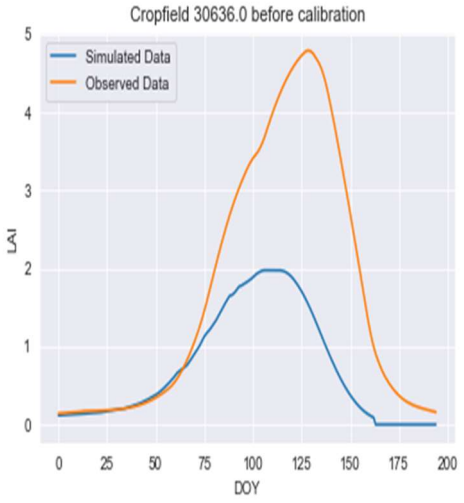

Cropfield 19099.0 after calibration
Model LAI
Observed LAI
LAI
DOY

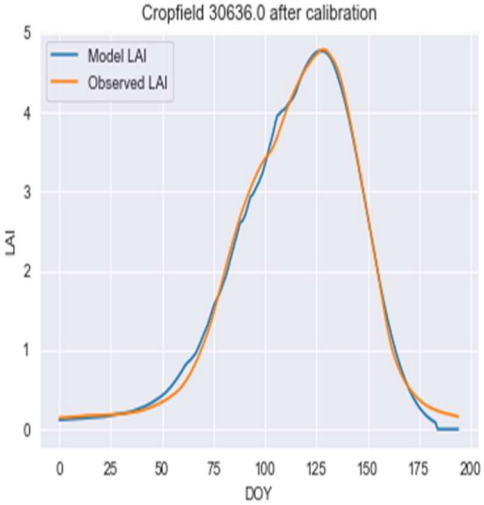

Cropfield 24572.0 before calibration
Simulated Data
Observed Data
LAI
DOY

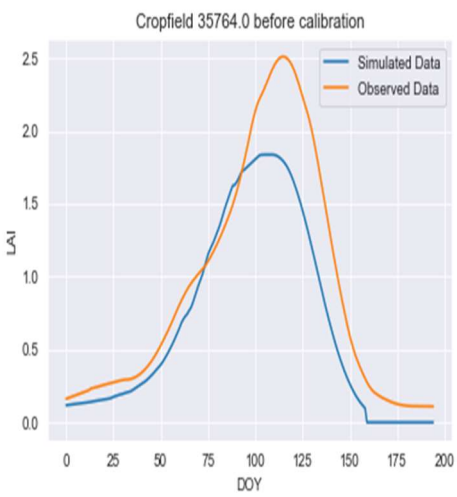

Cropfield 24572.0 after calibration
Model LAI
Observed LAI
LAI
DOY

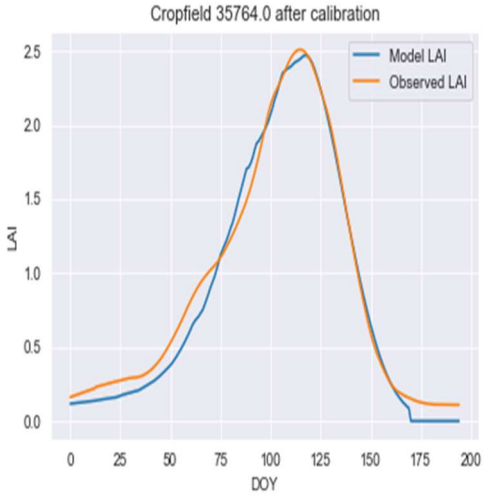

Cropfield 25261.0 before calibration
Simulated Data
Observed Data
LAI
DOY

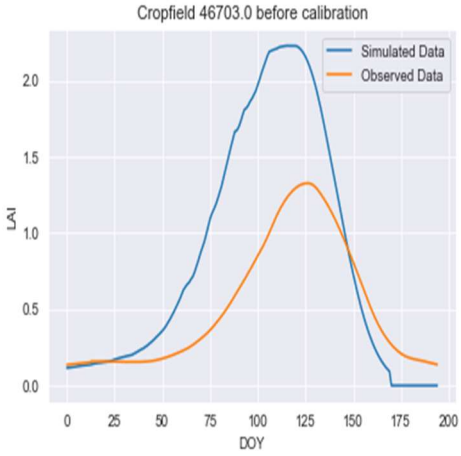

Cropfield 25261.0 after calibration
Model LAI
Observed LAI
LAI
DOY

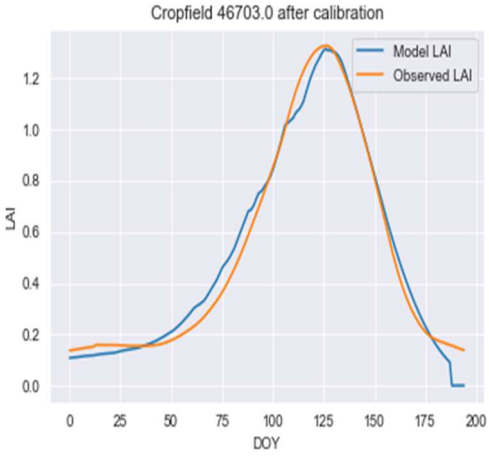

Cropfield 30636.0 before calibration
Simulated Data
Observed Data
LAI
DOY

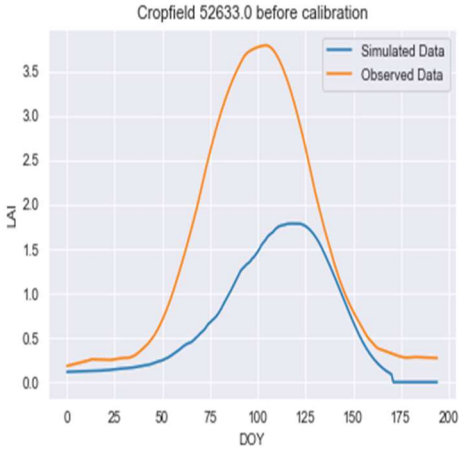

Cropfield 30636.0 after calibration
Model LAI
Observed LAI
LAI
DOY

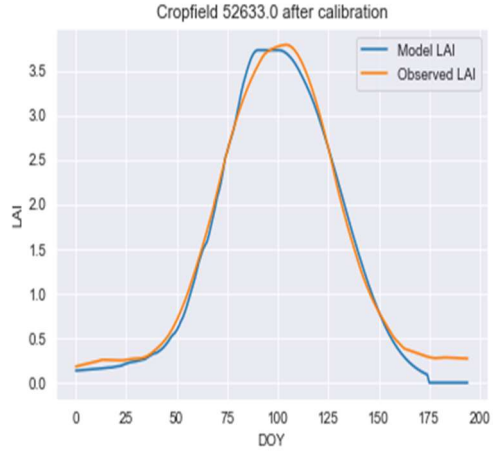

Cropfield 35764.0 before calibration
Simulated Data
Observed Data
LAI
DOY


Cropfield 35764.0 after calibration
Model LAI
Observed LAI
LAI
DOY

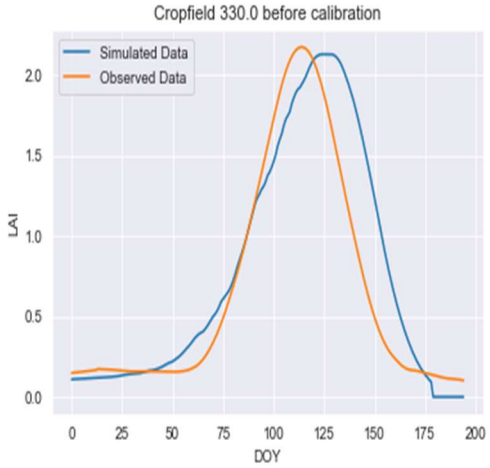

Cropfield 46703.0 before calibration
Simulated Data
Observed Data
LAI
DOY

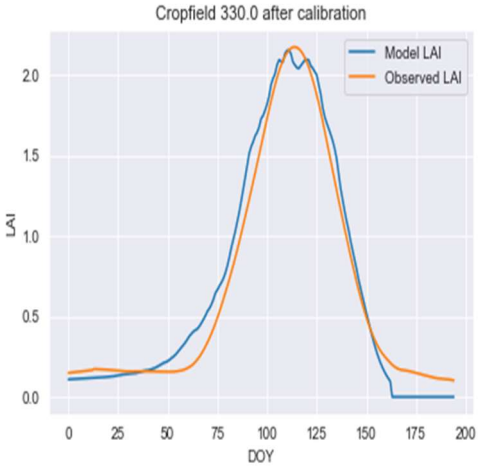

Cropfield 46703.0 after calibration
Model LAI
Observed LAI
LAI
DOY

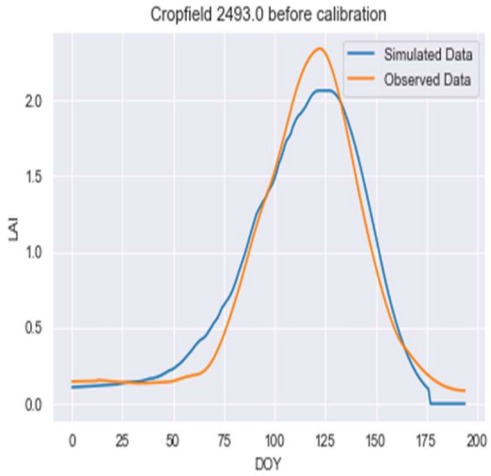

Cropfield 52633.0 before calibration
Simulated Data
Observed Data
LAI
DOY

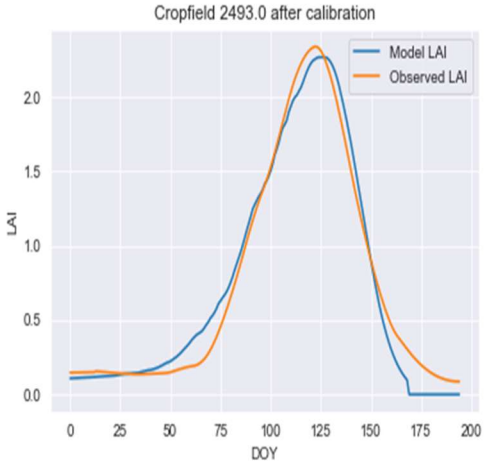

Cropfield 52633.0 after calibration
Model LAI
Observed LAI
LAI
DOY

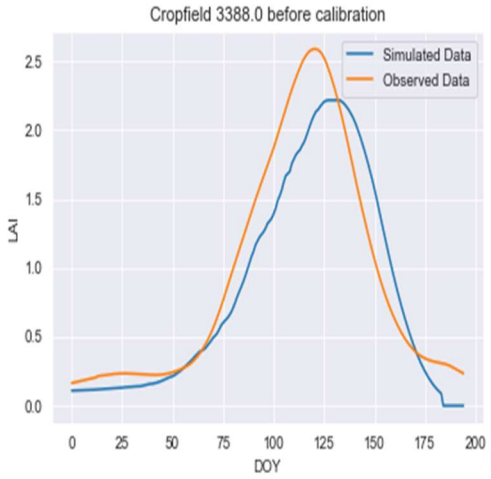
Appendix B: LAI simulation for selected cropfields using the best2bin DE method

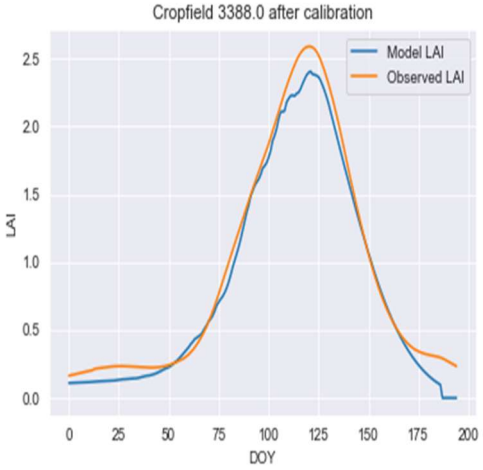


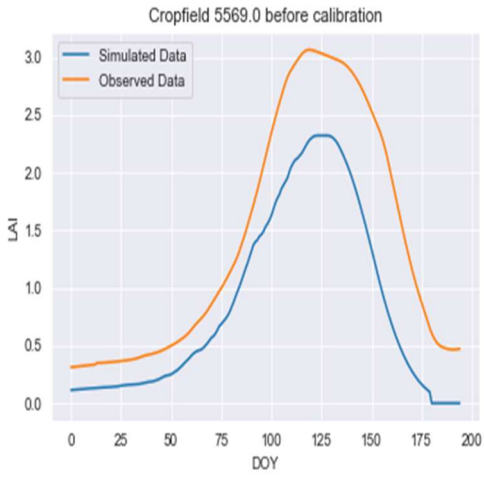


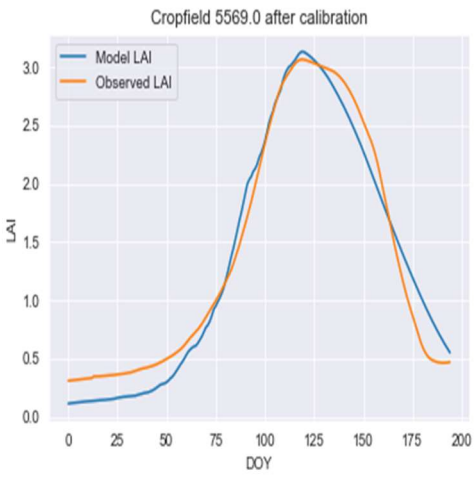


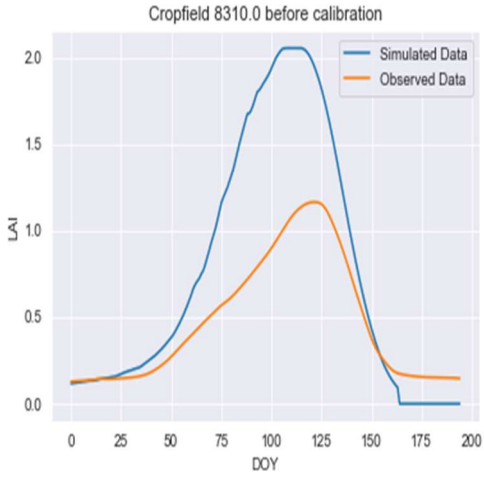


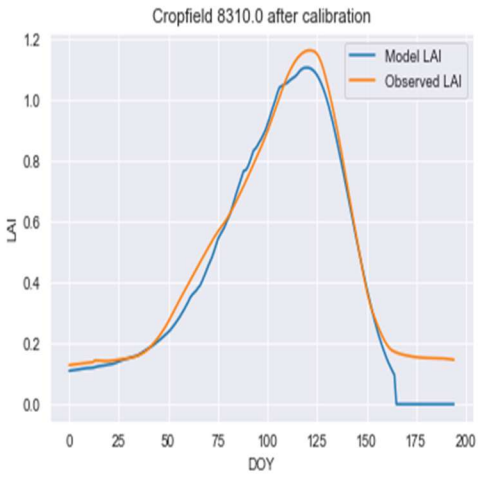


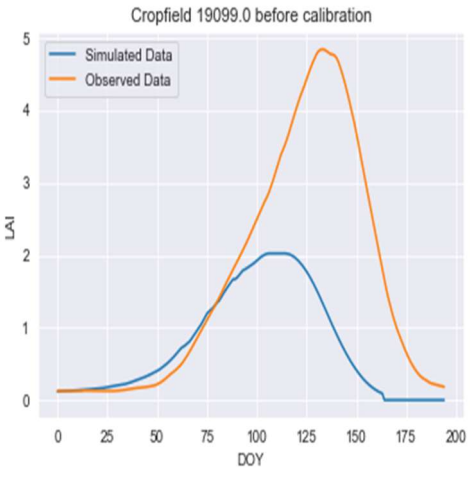


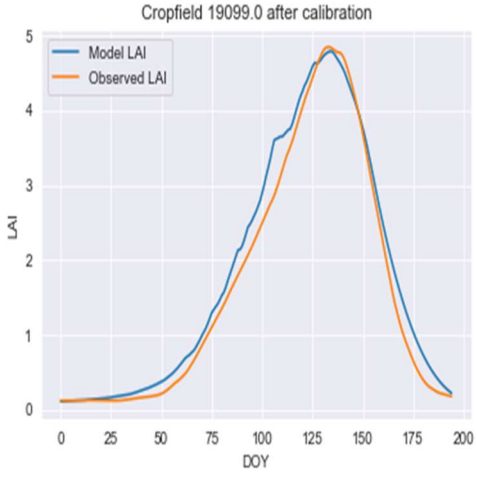

Cropfield 5569.0 before calibration
Simulated Data
Observed Data
LAI
DOY

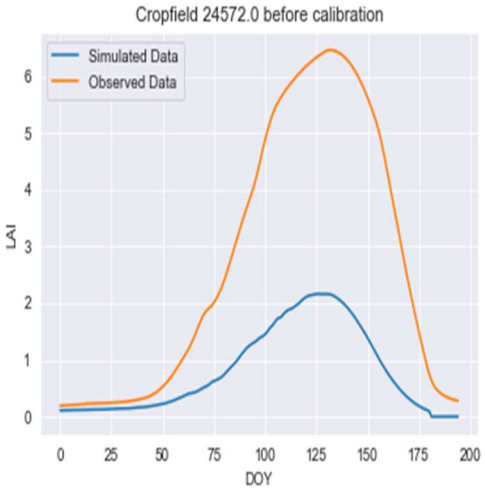

Cropfield 5569.0 after calibration
Model LAI
Observed LAI
LAI
DOY

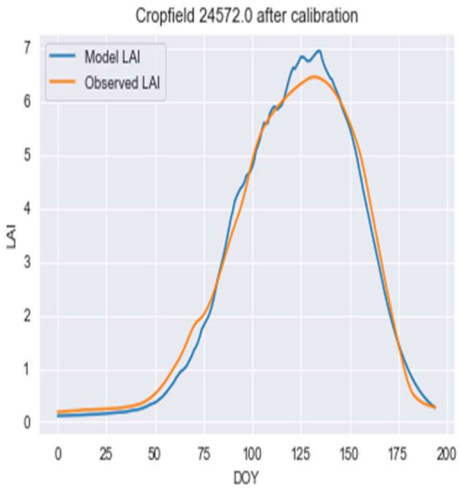

Cropfield 8310.0 before calibration
Simulated Data
Observed Data
LAI
DOY

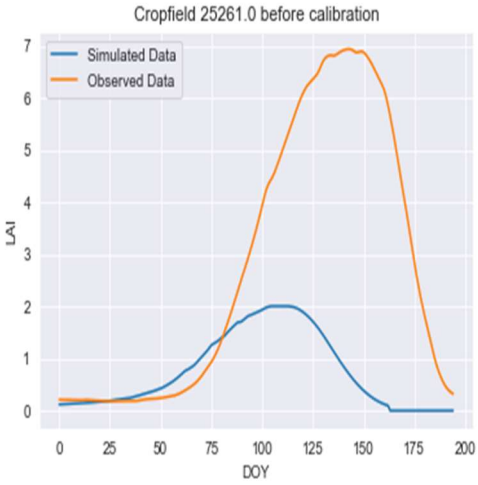

Cropfield 8310.0 after calibration
Model LAI
Observed LAI
LAI
DOY

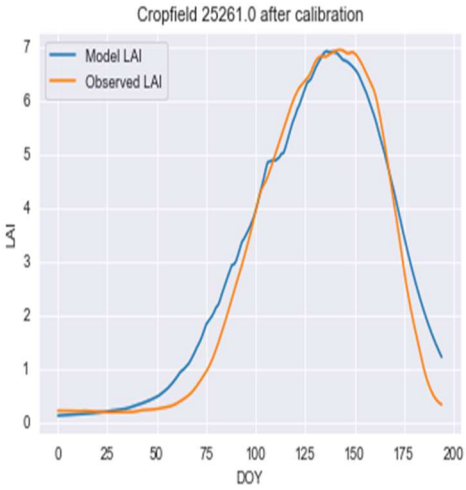

Cropfield 19099.0 before calibration
Simulated Data
Observed Data
LAI
DOY

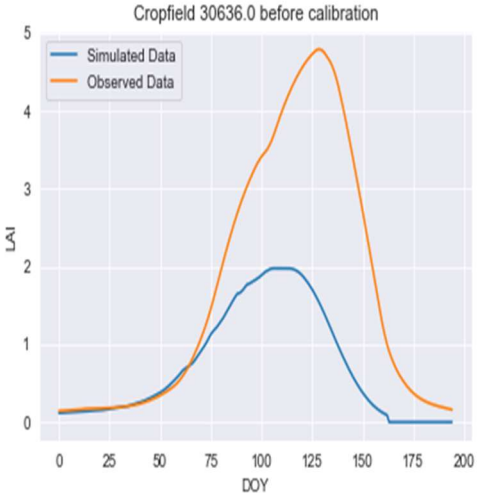

Cropfield 19099.0 after calibration
Model LAI
Observed LAI
LAI
DOY

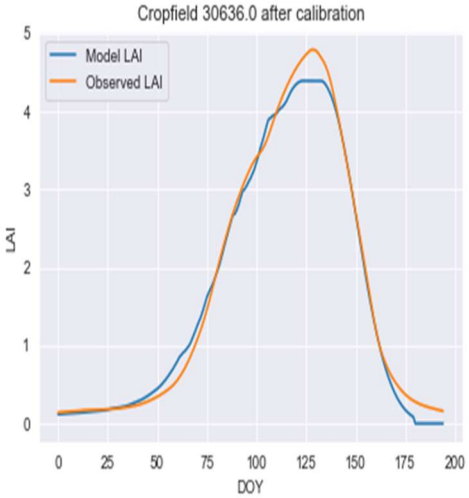

Cropfield 24572.0 before calibration
Simulated Data
Observed Data
LAI
DOY

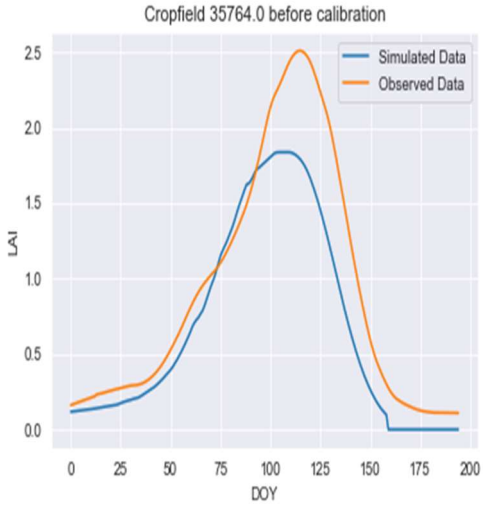

Cropfield 24572.0 after calibration
Model LAI
Observed LAI
LAI
DOY

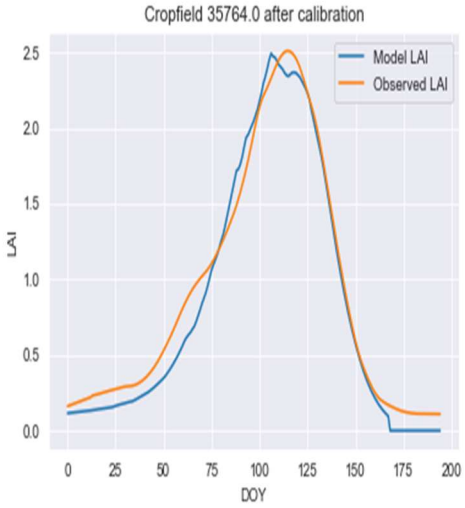

Cropfield 25261.0 before calibration
Simulated Data
Observed Data
LAI
DOY

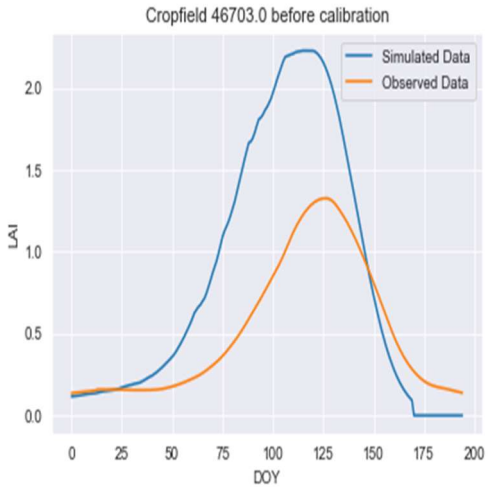

Cropfield 25261.0 after calibration
Model LAI
Observed LAI
LAI
DOY

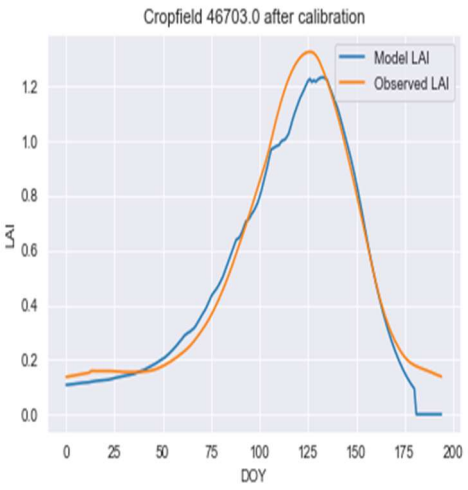

Cropfield 30636.0 before calibration
Simulated Data
Observed Data
LAI
DOY

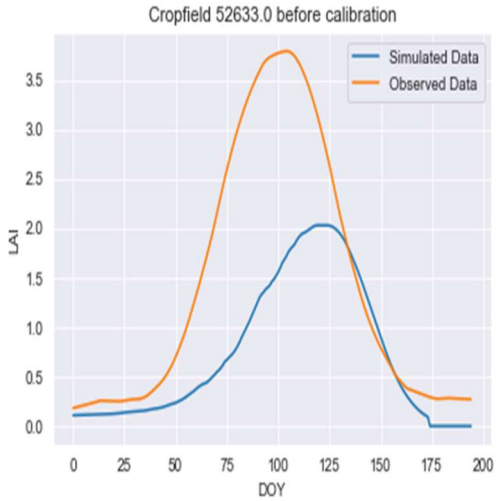

Cropfield 30636.0 after calibration
Model LAI
Observed LAI
LAI
DOY

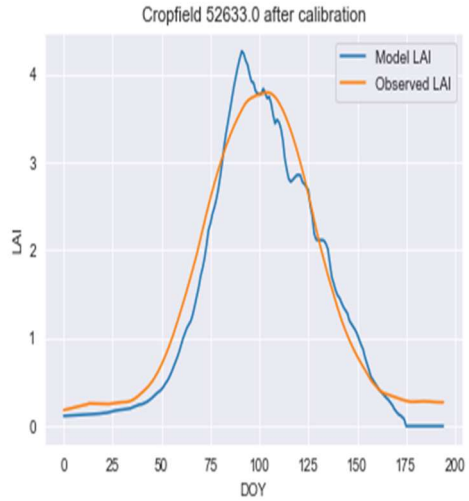

Cropfield 35764.0 before calibration
Simulated Data
Observed Data
LAI
DOY


Cropfield 35764.0 after calibration
Model LAI
Observed LAI
LAI
DOY


Cropfield 46703.0 before calibration
Simulated Data
Observed Data
LAI
DOY


Cropfield 46703.0 after calibration
Model LAI
Observed LAI
LAI
DOY


Cropfield 52633.0 before calibration
Simulated Data
Observed Data
LAI
DOY


Cropfield 52633.0 after calibration
Model LAI
Observed LAI
LAI
DOY

Appendix C: Wheat yield and maximum LAI obtained for each cropfield

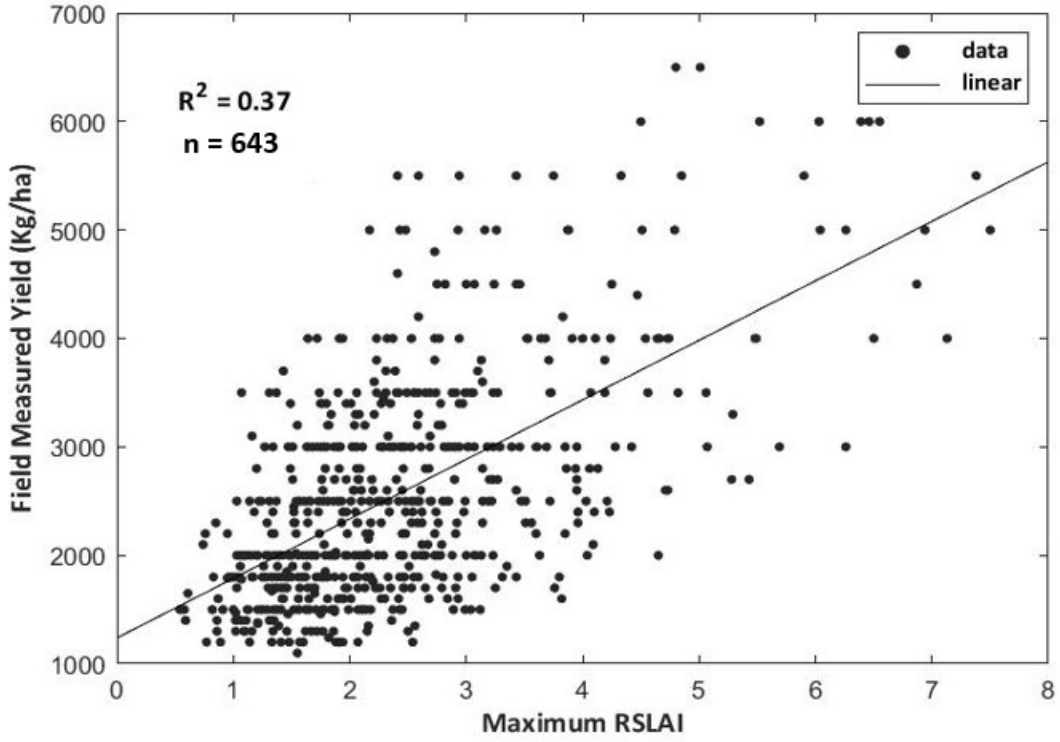


Figure C.1: Scatterplot of the relationship between all wheat yield data and maximum smoothed LAI.

Improving wheat crop yield estimation by assimilation of remote sensing biophysical variable in the Simple Algorithm for Yield Estimation (SAFY) model: A case study of Spain

Ibrahim Raufu

Reliable crop yield estimation is fundamental for effective agricultural management and food security. However, conventional methods such as crop-cutting trials and household surveys are expensive, time-consuming, and labour-intensive. In this master's thesis, we adopt the use of biophysical variables retrieved from high spatial and temporal resolution remote sensing data, particularly Sentinel-2 (S2), coupled with the Simple Algorithm for Yield estimation (SAFY) model to improve the estimation of wheat yield at the field level in the regions of Castile and Leon and Castile-La Mancha in Spain. Initially, we used meteorological time-series data to estimate wheat yield for rainfed, irrigated, and combined cropfields. However, our findings revealed that the model could not adequately capture the wheat phenology, leading to an underestimation of the yield. Due to the lack of correlation between the estimated and field-measured wheat yield, the Root Mean Square Error (RMSE) values of 1325 kg/ha, 3289 kg/ha, and 1722 kg/ha were obtained for the three model simulation cases. To address this limitation, we improved the model performance by assimilating S2 derived leaf area index (LAI) into the SAFY model to optimize six free parameters using the Differential Evolution (DE) algorithm. The results showed a moderate correlation and improved wheat yield estimation for rainfed (R^2 = 0.292 and RMSE = 744 kg/ha), irrigated (R^2 = 0.152 and RMSE = 1786 kg/ha), and combined cropfields (R^2 = 0.378 and RMSE = 952 kg/ha). Overall, the research results demonstrate the potential of high-resolution remote sensing data in conjunction with a simple crop model for improving crop yield estimation at the field level.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

Faculté des bioingénieurs

Croix du Sud, 2bte L7.05.01, 1348 Louvain-La-Neuve, Belgique | www.uclouvain.be/agro