Faculté des bioingénieurs

Assessing the Potential of Parcel-Level Barley Yield Estimation in Spain : A Case Study Using the Process-Based Model SAFYE-CO2 and the ACEO Processing Chain

Auteur : Lena Margarete Rester
Promoteur(s) : Prof. Pierre Defourny, Pierre Houdmont
Lecteur(s) : Dr. Sophie Bontemps, Prof. Xavier Draye
Année académique 2024-2025
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 express my sincere gratitude to all those who supported and guided me throughout the journey of this thesis.

I thank Professor Pierre Defourny for his time, mentorship, and encouragement. Our in-depth discussions helped steer me in the right direction and helped ensure that the work meets a high scientific standard.

I am equally grateful to Pierre Houdmont for his continuous support, technical expertise, and guidance throughout this project. His ability to explain complex concepts and share valuable insights into the field significantly shaped my understanding of the topic. I am especially grateful for the bi-weekly meetings he took the time to hold, even during particularly busy periods, which provided not only direction but also motivation. His consistent support ensured that I had the tools and resources needed to carry this project to completion.

My sincere thanks also go to Ahmad Al Bitar, Eric Ceschia, and Veronika Antonenko from the Centre d'Etude Spatiales de la Biosphère (CESBIO) for generously sharing the SAFYE-CO2 and ACEO model code and for providing crucial technical insights and support.

I would like to acknowledge the GEM programme and the financial support of the Erasmus Mundus Joint Master scholarship that enabled me to pursue this programme. Thanks to the dedicated lecturers from Lund University and UCLouvain I gained knowledge far beyond what is presented in this thesis. They opened my eyes to the potential of remote sensing and environmental modeling, and inspired me to reflect critically on the environmental and social dimensions of agriculture - lessons I will carry forward in my future work.

To my family and friends, thank you for your unwavering moral support throughout this journey. Your constant encouragement, patience, and belief in me helped carry me through moments of doubt and fatigue. Whether it was a kind word, a shared laugh, or a home-cooked meal, your presence - near or far - reminded me that I was never alone. I am especially grateful for the little things you did to keep me grounded and motivated; they meant more than you know. I am especially grateful to my fellow GEM colleagues: thank you for the camaraderie, shared struggles, endless technical discussions, and the humor that uplifted us during challenging times.

I would also like to sincerely thank my jury members, Dr. Sophie Bontemps and Prof. Xavier Draye, for dedicating their time and expertise to evaluate this thesis. I truly appreciate their careful reading and thoughtful feedback. It is my hope that they find the work both intellectually engaging and a meaningful contribution to the field.

Lastly, I am thankful for ChatGPT, which improved the clarity and accessibility of this work, and supported my learning journey in programming and scientific writing.

Table of Content

List of Abbreviations

ACEO	AgriCarbon-EO processing chain
APSIM	Agricultural Production Systems Simulator
BASALT	Bayesian Normalized Importance Sampling via LUT Generation
BO	Bayesian Optimization
CESBIO	Centre d'Etudes Spatiales de la Biosphère
CGM	Crop Growth Model
CV	Coefficient of Variation
CyL	Castilla y León
DAM	Dry Aboveground Biomass
DBM	Dry Below ground Biomass
DGM	Dry Grain Biomass
DLM	Dry Leaf Biomass
DOE	Day of Emergence
DOH	Day of Harvest
DOY	Day of Year
ERA	ECMWF ReAnalysis
ESDAC	European Soil Data Centre
ESYRCE	Encuesta sobre Superficies y Rendimientos de Cultivos
FAO	Food and Agricultural Organization
GEOGLAM	Group on Earth Observation Global Agricultural Monitoring
GLA	Green Leaf Area
GPP	Gross Primary Productivity
HI	Harvest Index
LAI	Leaf Area Index
L-BFGS-B	Limited memory Broyden–Fletcher–Goldfarb–Shanno algorithm with Bounds
LHS	Latin Hypercube Sampling
LUE	Light Use Efficiency
LUT	Look-Up Table
MAE	Mean Absolute Error
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetation Index
NPP	Net Primary Productivity
NSE	Nash-Sutcliffe Efficiency
NSO	National Statistical Office

PAR	Photosynthetically Active Radiation
RMSE	Root Mean Square Error
RRMSE	Relative Root Mean Square Error
RS	Remote Sensing
SAFY(E-CO2)	Simple Algorithm For Yield estimates
SCYM	Scalable Crop Yield Mapper
SDG	Sustainable Development Goal
SLA	Specific Leaf Area
SMT	Sum of Efficient Temperature since emergence
TOA	Top of Atmosphere
UAV	Unmanned Aerial Vehicle
US	United States of America
WOFOST	World Food Studies

List of Figures

List of Tables

1. Introduction

In 2015, the United Nations agreed to the 17 Sustainable Development Goals (SDGs), which aim to “end poverty, protect the planet, and ensure that by 2030 all people enjoy peace and prosperity” (United Nations Development Programme, 2022). The second goal, SDG 2: “Zero Hunger”, aims to “end hunger, achieve food security and improved nutrition and promote sustainable agriculture” (United Nations General Assembly, 2015). However, the progress toward achieving the goal by 2030 is too slow and as stated by the Food and Agriculture Organization (FAO): “The world is not on track to achieve targets for any of the nutrition indicators by 2030” (FAO et al., 2021, p. 38).

Human-induced natural disasters, political conflict and war, a growing population, and the decline in soil quality are few but very important causes of the continuous and increasing global food insecurity (FAO, 2017; IPCC, 2018). The global food crisis between 2008 and 2011 is a recent global example of how food price speculations, crop failures of major grain- and cereal production, as well as distortions in international markets have driven millions of people into poverty and food insecurity (United Nations, 2011). This disaster made clear that timely and accurate monitoring of food production is needed to decrease the risk of future price volatility in global food markets (Fritz et al., 2019).

As a response, the Group on Earth Observation (GEO) initiated the GEO Global Agricultural Monitoring (GEOGLAM) which was designed to provide monthly global assessments and predictions of food production of major crops (Fritz et al., 2019). It makes use of recent technological advances in both modelling and remote sensing (RS) technologies. Remote sensing technologies have seen a rise in applications, due to advances in their temporal, spectral, and spatial capabilities (Weiss et al., 2020). It is in particular interesting for food production due to its non-destructive global coverage of land surfaces which provides important insights into crop development, being an alternative to time and cost intensive manual surveys (Basso & Liu, 2019).

Nonetheless, while timely assessments and forecasts of crop production exist, such as from GEOGLAM, they are limited in their information at high spatial resolutions, such as providing sub-national to parcel level information. This information is of particular interest in highly fragmented and diverse landscapes such as in Europe or countries with high food insecurity prevalence, for example in Africa and Southeast Asia (Basso & Liu, 2019). It can help farmers and local economies make informed decisions about prices, storage needs, and farm management (Basso & Liu, 2019). Additionally, it enhances understanding of spatial differences in crop productivity, supporting better planning, especially under shifting climatic conditions that affect local crop suitability (Abd-Elmabod et al., 2020). Aggregated, the information can be used at national scales to inform import and export strategies, and regulate food prices (Fritz et al., 2019).

Therefore, the primary goal of this study is to support research in crop yield production at high spatial resolution by making use of recent advances in process-based modelling for crop growth modelling and remote sensing. It is applied to the highly fragmented and diverse landscape of Castilla y León (CyL) in Spain for barley production. In addition, attention will be paid to the need of forecasts and comparison to other statistical methodologies that can be used for estimating crop yields.

An in-depth literature review is presented in Section 2, covering the background of the study area and barley cultivation, current yield estimation methodologies, and the role of remote sensing in this context. This review helps to frame and motivate the research objectives outlined in Section 3. The research objectives are addressed in the Results, Section 4, and contextualized in the Discussion, Section 5, which highlights both the benefits and challenges of the chosen approach. Finally, the paper concludes with a summary of key findings and recommendations for future research in Section 6.

2. Literature Review

2.1 Barley production in Spain

Barley has been cultivated in Spain for several millennia, but its production significantly increased during the 20th century (Martínez-Moreno et al., 2024). In 1975, for the first time, the area dedicated to barley cultivation surpassed that of wheat, and barley has remained the most widely cultivated cereal in Spain for the subsequent 45 years (Martínez-Moreno et al., 2024). At present, the area allocated to barley is approximately 2.3 to 2.5 million hectares, with an average annual harvested production of 8 million tons (Martínez-Moreno et al., 2024). This accounts for roughly 10–20% of total barley production in Europe, positioning Spain as the third-largest barley producer in the European Union, following Germany and France (Eurostat, 2025). In 2020, 32.3% of the total barley production area in Spain was located in Castilla y León (Ministerio de Agricultura Pesca y Alimentación, 2020). The main variety of barley produced is two-row winter barley (Ministerio de Agricultura Pesca y Alimentación, 2020). The two-row categorization is based on the barley's spike morphology. It is considered a winter crop due to its sowing time during the winter months; see Table 1 (Piferrer et al., 2021).

Table 1: Distribution of sowing and harvest times across the months for Castilla y León between 2014-2016 for 2- row Barley based on an ESYRCE survey (Piferrer et al., 2021).

	No. of parcels	Percentage of parcels per month											
		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Sowing	757,378	6	6	1							22	47	18
Harvest	2,431,157						20	64	16				

The recent expansion in barley cultivation is closely linked to its primary use: as animal feed. As living standards in Spain improved throughout the 20th century, meat consumption increased, leading to a greater demand for feed grains such as barley (Martínez-Moreno et al., 2024). In addition to its grain being used for feed, barley straw is commonly used as livestock bedding. Smaller quantities of the grain are also used for human consumption and, since the mid-20th century, in beer production (Martínez-Moreno et al., 2024). By 2022, Spain had become the second-largest beer producer in the EU, surpassed only by Germany. However, the proportion of barley used for brewing remains relatively low, with approximately 8% of the cultivated area dedicated to malting barley and the remaining 92% used for feed (Martínez-Moreno et al., 2024). Martínez-Moreno et al. (2024) project that in the future, about two-thirds of Spain's barley production will continue to be used for animal feed, while one-third will serve brewing and food purposes.

Given barley's economic importance—closely tied to both the livestock and brewing industries—timely and accurate assessments of seasonal barley production are increasingly essential, particularly ahead of the harvest. In light of the predicted climate changes affecting Spain, yield estimation methods that can account for meteorological conditions will be critical tools for improving yield forecasts. Over the past 100 years, i.e., from 1901 to 2023, the observed annual mean surface air temperature significantly increased from 10.5°C to 13.2°C (Climate Change Knowledge Portal, n.d.), and is predicted to increase by 1.3 – 2.1°C between 2030 and 2050, compared to the baseline of 1971-2000 (Fernández et al., 2019). At the same time, Senent-Aparicio et al. (2023) found that the number of rainy days will decrease particularly in March and June, yet with a simultaneous increase in total precipitation, causing extreme precipitation events. In this context, Cammarano et al. (2019) show that an increase in dry and hot days, particularly during the growing season of barley, and before the sowing, will negatively impact barley yield: a dry scenario could decrease the yield by 25-50%. An increase in precipitation, however, could increase yields by 25-50%.

To identify suitable approaches for yield estimation in this context, the following section introduces methods for acquiring yield data through survey sampling. It then presents the two main categories of yield estimation methods, i.e., statistical and process-based models, as well as how remote sensing products can support both approaches.

2.2 Yield survey sampling

Survey sampling methods are used to generate ground-truth or in-situ crop yield data. Crop yield is defined as "the ratio of the total mass of harvested product (e.g., a grain) to the area used to grow the crop" (Weiss et al., 2020), meaning the average quantity of crop produced per unit area. This metric is used to calculate the total crop production at various scales, i.e., from parcel to farm to nation, which reflects the quantity of crop harvested from the area used for cultivation (Organisation for Economic Co-Operation and Development, n.d.). The following section outlines the main survey sampling approaches as well as their limitations and error sources, based on the summary by Sapkota et al. (2016).

Crop cuts have long been the standard method. With this method, the yield in one or more sub-plots, often smaller than $10m^2$, is measured by cutting and weighing the grain. To achieve representativeness of the field's heterogeneity and avoid over- or underestimation, multiple sub-plots need to be cut. This makes it a very destructive method, requiring farmer agreement and compensation. Similar to *crop cuts* yet less destructive is the *grain weight/test weight* method which counts the average number of ear-heads/pods within multiple sub-plots and relates this information to the yield. The advantage of both methods is that they can be performed before the end of the season.

In contrast, the *whole plot harvest* method is conducted by measuring the yield directly at the time of harvest. It is often considered as the most accurate method, Lobell et al. (2020) also term it the "gold standard". Yet, a condition for this method is that the crop is not harvested at different moments, resulting in a high labor demand for a short peak period.

A non-destructive method and very commonly used is the *farmer survey* method. Farmers are interviewed and asked to estimate (before the harvest) or recall (after the harvest) their yield. While Murphy et al. (1991) identify farmers' estimates as highly accurate, Lobell et al. (2020) point out that they often contain systematic errors. These can be caused by the recall bias, the tendency to round off numbers, the utilization of unstandardized measurement units, imprecise information regarding the crop conditions and harvest times (Carletto et al., 2015), or the incentive to purposefully (and sometimes sub-consciously) under- or overestimate their yield due to taxes (Diskin, 1999).

Lastly, the *expert assessment method* relies on the expertise of agronomists, extension agents, or researchers who visually assess the field. This avoids the subjectivity of farmer-reported data, but the accuracy depends highly on the level of expertise of the assessor (Sapkota et al., 2016).

While survey sampling methods have a high time and labor demand, they are needed to generate ground-truth data. They are the standard method employed by National Statistical Offices (NSOs) (European Parliament Council of the European Union, 2021; Sapkota et al., 2016), to generate information on the country's crop production. The method implemented is highly dependent on the country's resources since methods like whole plot harvest and expert eye assessment require trained people at a certain time of the year. In Spain, expert eye assessment is the primary method, whereas African countries mainly rely on farmer reports as part of household surveys such as the World Bank's "Living Standards Measurement Study Integrated Survey on Agriculture" (Lobell et al., 2020). This leads to a high variability across countries in their (continuous) implementation and therefore the comparability and standard of quality (López-Lozano et al., 2015) since each method introduces its own source of errors, as outlined above.

To improve reliability and consistency across national yield estimation reports as well as to extrapolate yield estimates, statistical and process-based models are being developed to leverage survey sampling data as calibration and validation inputs (Basso & Liu, 2019; Sapkota et al., 2016; Weiss et al., 2020).

2.3 Yield estimation methods

2.3.1 Statistical models

Statistical models, an empirical method, calculate a regression between agrometeorological variables and (historical) yield data from survey sampling. Then, using agrometeorological data as inputs, the established regression can estimate yield for areas where no survey sampling took place or forecast yield prior to the harvest. The agrometeorological inputs can be weather data, vegetation indices, soil characteristics, or agronomic variables such as the sowing date (Basso & Liu, 2019). Commonly used algorithms include linear regressions, traditional machine learning, and, considering new technological advances, more recent deep learning methods.

The advantage is that statistical models are relatively simple and need few input parameters (Basso & Liu, 2019). If trained on a specific area, time and crop, they produce accurate results (see examples: Cao et al., 2021; Paudel et al., 2021).

The disadvantage is that for proper training, sufficiently large historical yield datasets are required. These can only be acquired using above-mentioned survey sampling methods which have been shown to have different levels of accuracy and are commonly concentrated in commercial systems and recent years (Deines et al., 2021).

Furthermore, even if properly trained, statistical models still suffer from limited spatial and temporal applicability, since regression models cannot extrapolate results outside of the boundaries of the data used for training (Deines et al., 2021; Lobell et al., 2015). Considering the increasing climate variability and occurrence of extreme events, it is challenging to apply statistical models for yield estimation for unseen climate events, especially when training datasets have limited climatic variability. Furthermore, spatial applicability is limited as the regression does not account for climate-soil-plant-management dynamics. Hence, in areas where the climate or management practices are different or a different crop is planted, the statistical model might no longer be accurate (Lobell, 2010).

An operational example of a statistical yield estimation model is the Sen4Stat toolbox. It extracts 43 phenological features from Sentinel-1 and Sentinel-2 time-series imagery, incorporates meteorological data from the ECMWF ReAnalysis 5 (ERA5) Land dataset, and relates these inputs to observed ground-truth yield data to train statistical models (Bontemps et al., 2024). Designed primarily for use by NSOs, Sen4Stat provides a scalable yield estimation approach if sufficient and high quality ground-truth data for training is available.

Another example of a large-scale application of a statistical model for yield estimation is the Integrated Canadian Crop Yield Forecaster (Chipanshi et al., 2015). It uses a statistical algorithm to relate meteorological data, remotely sensed vegetation indices, crop and soil information to yield at census agricultural regions. Historical yield surveys of farmers are used as training data. While the model is operational, Chipanshi et al. (2015) have identified limitations of this method that reflect the aforementioned challenges of statistical methods. Firstly, Chipanshi et al. (2015) challenge the quality of the training data, as the reporting method is vulnerable to errors, the frequency in samples is low and the willingness of farmers to reply is decreasing, leading to poor-quality training and validation data. Secondly, the model is not sensitive to the extreme ends of the yield spectrum as it overestimates low yields and underestimates high yields (Chipanshi et al., 2015). This reflects the temporal applicability challenge of statistical methods, as they are trained on historical data and hence cannot account for unseen extreme events.

In summary, while statistical models might require less input data and can produce highly accurate results for a limited temporal and spatial framework, the quality issues of survey sampling data are propagated and statistical models require continuous adjustments to reflect changes in the environment. Considering the climatic challenges expected for Spain's barley production, statistical methods appear unfit. To avoid these issues, process-based models, while not without their own flaws, provide an alternative.

2.3.2 Process-based models

As the name suggests, process-based models aim to quantify a complex reality in the form of mathematical equations and thereby simulate the different processes and relationships within the agricultural system (Basso & Liu, 2019). They use meteorological data, soil characteristics, crop, and management information as inputs to mathematical equations which simulate plant development and growth. In comparison to statistical models, process-based models do not aim at finding a regression between agrometeorological variables and yield observations, but rather account for the physical and biological interactions taking place on a field, hence in the agricultural science domain they are often termed *crop growth models* (CGM) (Basso & Liu, 2019; Weiss et al., 2020).

Luo et al. (2023) divide existing CGMs into four categories, based on their core principle of plant growth simulation. There are *light driven* models, such as SAFY (Simple Algorithm For Yield estimates), where the relationship between light energy utilization and crop dry matter accumulation is the main driver for crop growth. *Soil-moisture driven* models, such as AquaCrop, focus on soil moisture as the main factor of crop growth. Thirdly, *atmospheric factor driven* models, such as WOFOST (WOrld Food Studies), describe the effects of atmospheric properties, e.g., CO2 availability, on crop growth. Lastly, *integrated factor driven* models integrate multiple sub-modules on biophysical and agronomic processes to simulate crop growth, encompassing many factors and synergies taking place at field and farm scale, such as the Agricultural Production Systems sIMulator (APSIM) (Holzworth et al., 2014; Luo et al., 2023).

The WOFOST model is one of the older models which has been developed by the Centre for World Food Studies in Wageningen, the Netherlands for modelling agriculture in the tropics with a spatial focus on national and regional planning (van Diepen et al., 1989). Growth and production of annual crops are simulated from emergence to maturity in daily steps and are based on CO_2 assimilation calculated from the absorbed photosynthetically active radiation (APAR) and the available leaf area (de Wit et al., 2019). Additional factors included in the processes of crop growth simulation are the crop species, soil type, hydrological conditions and the weather (van Diepen et al., 1989). The yield is estimated through partitioning of the accumulated biomass into storage organs. The main stress factors for crop development were initially related to light, temperature, water and nutrient. However, de Wit et al. (2019) outline the improvements made to the WOFOST model over its more than 25 years of operational use: next to the initial stress factors, the model has been expanded to include crop development inhibition caused by weeds, pests, diseases, or pollutants. Overall, the model is considered to be rather complex, including numerous processes which require a large number of parameters, to be provided by the user of the model. This is a challenge in using WOFOST as the parameters can be unavailable and difficult to measure (Boogaard et al., 2013). Furthermore, de Wit et al. (2019) highlight that the model results are not suitable for small scales, but should preferably be aggregated to grid and regional levels.

AquaCrop on the other hand uses a small number of parameters, to achieve a balance of simplicity, accuracy, and robustness, as described by Steduto et al. (2009). It was developed by the FAO and is designed for end-users and practitioners. In contrast to other models, AquaCrop is water instead of light driven, meaning that crop growth is considered a function of water consumption. In AquaCrop, first transpiration is calculated from which biomass is inferred on daily time steps, using a crop-specific biomass-water-productivity parameter. In its calculations, the model uses inputs on the weather and soil characteristics, however influences of pest and diseases are not considered. Eventually, crop yield is a fraction of the total biomass, determined a harvest index. While the model requires much fewer inputs from the user, a water-driven approach is still complex as the responses of crops to water stress are not yet fully understood and difficult to model as the effect depends on the intensity, duration, and time of occurrence of the water stress (Steduto et al., 2009).

As in the name, SAFY is a simple process-based model, where crop growth is driven by a light-use-efficiency (LUE) approach as developed by Monteith (1977). It linearly relates intercepted radiation to produced dry above-ground mass (DAM) (Duchemin et al., 2008). The crop growth is calculated daily from plant emergence to the day of completed leaf senescence based on the leaf area of a plant (Duchemin et al., 2008). The leaves absorb PAR and increase the plant's DAM, which in turn is partitioned partly into more leaf area for photosynthesis. The LUE thereby controls the DAM production by accounting for multiple plant stress factors such as water and temperature stress. Grain yield is derived as a proportion of total DAM (Duchemin et al., 2008). Over the years, the model has been improved to account for more biophysical processes, improving the biological realism of the model, yet increasing the number of inputs and parameters. Duchemin et al. (2015) have added a water balance module creating SAFY-WB, Pique et al. (2020a) added a carbon module to quantify the carbon balance creating SAFY-CO2, as well as combining both versions to create SAFYE-CO2 (Pique et al., 2020b). Instead of directly relating LUE to DAM, the latest models first calculate gross primary productivity, then autotrophic respiration of the plant, to eventually allocate the biomass to different plant organs, such as roots, leaves, and the grain. Furthermore, temperature and water stress are not only accounted for through the LUE, but are directly calculated within the model. At a minimum, the model requires daily weather data, as well as soil and vegetation parameters which have been made available for certain crop types by the model developers (Al Bitar et al., 2023). Additional information that can be included are management practices regarding fertilization and irrigation. Consequently, the latest version of the SAFY-models strikes a good balance between a realistic representation of biophysical processes and simplicity with the aim of being used at field scale.

On the other end of the spectrum of complexity are agro-econometric models such as APSIM. These models are designed to encompass multiple sub-models responsible for different processes, e.g., photosynthesis, water balance, or farm management, to not only model crop growth and yield, but rather assess (long-term) environmental impacts (e.g., climate risks, different crop types, presence of livestock, soil erosion and degradation) and management impacts (e.g., irrigation, tillage, crop rotation) on the farm ecosystem as a whole (Holzworth et al., 2014; Keating et al., 2003). APSIM has been developed in 1990 (and since continuously updated and improved) by the Agricultural Production Systems Research Units in Australia, with a focus on understanding and modelling the effects of climate change and risk on economic and ecological outcomes of management practices, including long-term effects (Keating et al., 2003). On the one side, the model consists of biophysical modules, which simulated plants (including specific species of crops, pastures, and forests, as well as mixed species systems and intercropping), soil, animals and the climate (Holzworth et al., 2014; Keating et al., 2003). For the biophysical models to work, data sets are required providing for example daily weather data, as well as information on specific parameters for site, cultivar, and management characteristics. On the other side, APSIM includes a management module, through which users can simulate farmer's decision-making, such as for example sowing, harvesting or killing crops, applying fertilizer, irrigation, or tillage (Keating et al., 2003). All models can be coupled as needed and communicate with each other to simulate the farm system as a whole (Holzworth et al., 2014; Keating et al., 2003). In Australia, the APSIM model has evolved into a Decision Support System for rainfed grain crop regions, which are used by consultants and farmers to analyze different management options (Holzworth et al., 2014).

While all of the process-based models do account for many soil-plant-atmosphere-management interactions, they are still a simplification of reality as they cannot capture the full complexity of reality (Weiss et al., 2020). Nonetheless, process-based models evade the scalability limitation of statistical models: since process-based models consider soil-plant-atmosphere-management dynamics, they can be applied to different locations and growing conditions. However, this does not mean that one model can simply be applied to each crop type and location. The model parameters need to be adjusted: vegetation parameters need to be representative of the crop type analyzed and soil parameters need to reflect the area's soil conditions. Especially in regional applications where input parameters have high spatial heterogeneity, the application of process-based models is limited if information on input parameters is not available (Luo et al., 2023). Such region- and crop-specific model parameters can either be derived through experiments and in-field measurements (for example, Pique et al., 2020a derived LUEb parameter from in-situ experiments across different countries) or through *model calibration*, which is the adjustment of model parameters to achieve model outputs that are in line with observations (Basso & Liu, 2019). A common method is to use remote sensing observations for model calibration as they capture "real-time observations" of crop status information over large areas (Luo et al., 2023). The combination of remote sensing information and process-based models is termed *data assimilation* and has gained increasing attention for the improvement of yield estimation accuracies of process-based models (Luo et al., 2023; J. Wang et al., 2024).

Therefore, in the following section data assimilation methods and their applicability to this research are described, identifying the best approach for model calibration. This is followed by an analysis of different model applications in previous literature to assess the performance of the SAFY models in comparison to other process-based models. Finally, the last section discusses the issue of operationalization and scalability of process-based models, highlighting the advances made with regard to the SAFYE-CO2 model.

2.4 Data Assimilation Methods

Luo et al. (2023) identify three main types of data assimilation methods. The most common method is *sequential assimilation* (or model-state-update-method) (Luo et al., 2023; J. Wang et al., 2024) whereby RS data is integrated stepwise into a crop model, updating the model's state variable(s), see Figure 1A. Algorithms, such as Kalman Filter or Particle filter, weigh the model estimate of the state variable with the remote sensing observations. An updated value of the state variable is then used as the starting point for the next simulation step until the next RS observation.

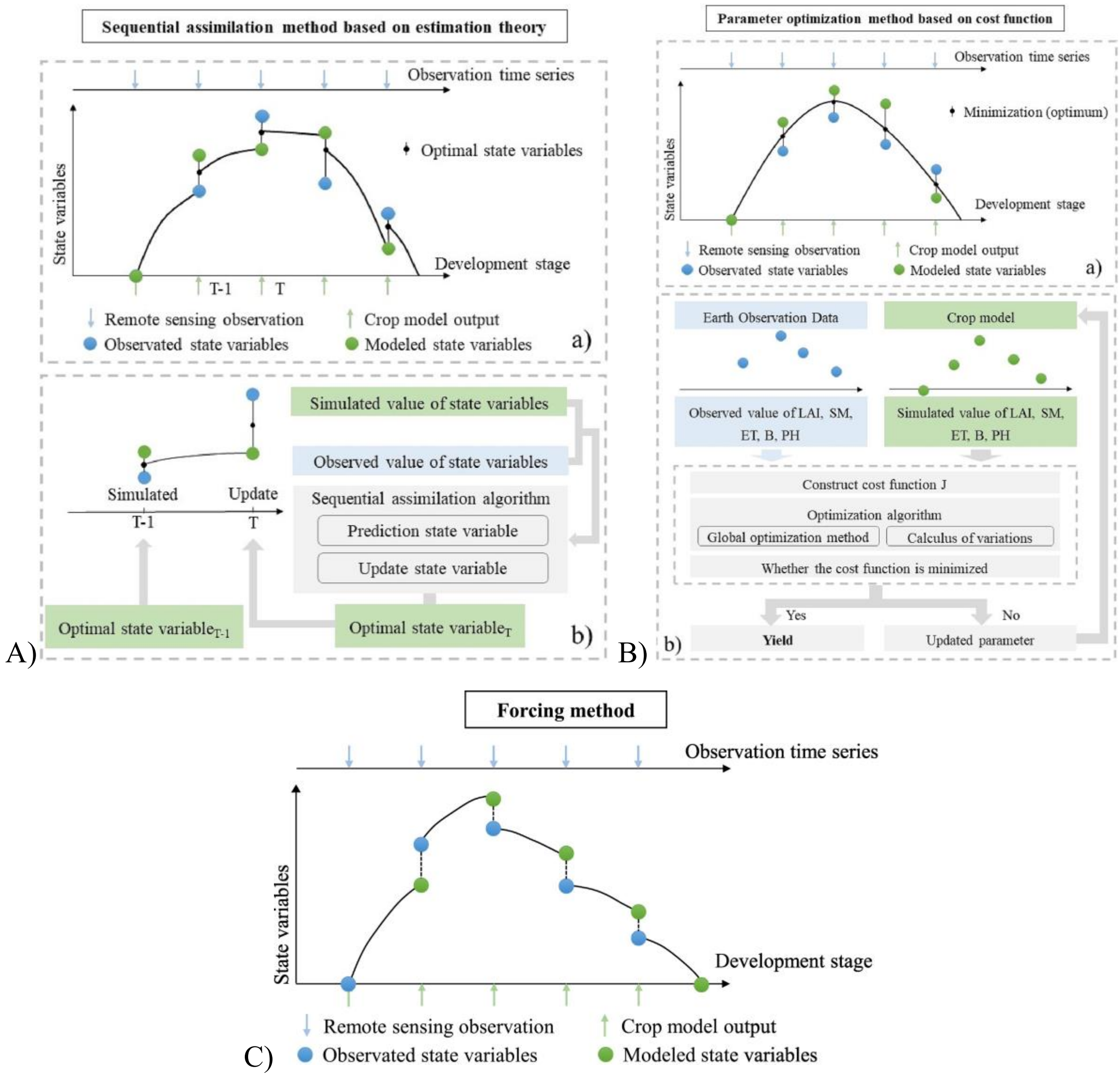


Figure 1: A) Sequential assimilation method illustration; B) Parameter optimization assimilation method illustration; C) Forcing assimilation method illustration (by Luo et al. (2023))

The second most common assimilation method according to Luo et al. (2023) is *parameter optimization (or calibration)*, whereby a modelled state variable is compared to the remotely measured observation. A cost function is used to adjust and optimize the model's parameters to minimize the difference between the modelled value of the state variable and the remotely sensed observation (see Figure 1B). Cost functions can be based on the Root Mean Square Error (RMSE), Least Square Method, or three- and four-dimensional variational methods. The result of a parameter optimization assimilation method is an optimized set of model parameters which will achieve the most accurate yield for the simulated conditions.

Lastly, the *forcing method* is a simple but less common method in which simulated values are directly replaced with RS observations (Luo et al., 2023) (see Figure 1C). While it is easy to compute and operate, the drawback of this method is that its accuracy depends on the quality of the RS observations, as it does not account for uncertainties in the RS data or the modelled values. Instead, errors in the observation data are carried over into the model simulation, resulting in lower accuracies, compared with the other two methods. Parameter optimization methods on the other hand yield better accuracies as the algorithms reduce errors from RS data, model initialization, or poor parameter choices (Luo et al., 2023). The iterative process, however, requires high computational costs and time, which is why sequential analysis is often a preferred method, although it does not provide an optimized set of parameters (Luo et al., 2023). However, the SAFY models have been designed to account for spatial variation between parcels through parcel-specific parameterization. This means that a certain number of parameters within the SAFY models should be adjusted to reflect the local conditions. Hence, parameter optimization is the most suitable method for this model and applied in this study.

Implementing the data assimilation process typically follows four steps, as outlined by Luo et al. (2023). First, a chosen crop growth model is calibrated to the use case, using previous literature or in-field measurements. Using this model, the state variables will be simulated. Secondly, RS measurements obtain spectral reflectance properties of the study area. The reflectance values are inverted into biophysical variables. The most widely used biophysical variable is the Leaf Area Index (LAI), which is defined as "half the developed area of photosynthetically active elements of the vegetation per unit horizontal ground area" (Deffense et al., 2021). Hence, it represents the amount of green leaves which are available for photosynthesis. Therefore, the LAI indirectly contains information on the effects of environment and management on crop growth and making it desirable for most models to estimate photosynthesis, biomass, and yield (Luo et al., 2023; J. Wang et al., 2024). Additionally, soil moisture is a key variable for understanding water availability and drought and is often combined with LAI for more robust results (Luo et al., 2023). Other variables that can be measured through remote sensing and used for assimilation include vegetation indices, e.g., NDVI (Normalized Difference Vegetation Index) and Enhanced Vegetation Index, or plant height (Luo et al., 2023). However, vegetation indices such as the NDVI saturate at LAI values of 2 to 6 (Baret & Guyot, 1991), making them less effective for the use of biomass estimations. In the third step, the obtained biophysical variables or vegetation indices are integrated into the crop model using an assimilation algorithm outlined above. Lastly, the assimilation results are validated using in-field measurements or official statistics, to assess the accuracy of the model output.

Overall, analyzing 143 studies, Luo et al. (2023) find that by using data assimilation methods, yield estimation accuracies improve. From the 143 studies, if model simulations were performed without data assimilation, the R^2 values range between 0.01-0.471. After data assimilation, R^2 values improved to 0.35-0.82. Moreover, the most common variable used for assimilation is the LAI (Luo et al., 2023; J. Wang et al., 2024). Therefore, to optimize the performance of SAFYE-CO2 for barley in CyL, calibration of location-dependent parameters through remotely sensed LAI, a crucial variable within the model, is applied in this research.

2.5 Different satellites for different purposes

An important factor in the success of the assimilation process is the quality and quantity of the remotely sensed data.

Quality refers to the spatial resolution of the satellite imagery, the accuracy of data processing, and the retrieval of biophysical variables from it. Until 2017, the spatial resolution of civil satellite imagery was very coarse, favoring its application in homogenous and large landscapes, such as in the United States (US) and China (Weiss et al., 2020). In contrast, Europe has heterogeneous landscapes due to a high diversity in crop types, crop management practices and field sizes (Duveiller & Defourny, 2010; Weiss et al., 2020). Similarly, African small-holder systems, especially areas threatened by food security, feature very small field sizes (Weiss et al., 2020). Such areas require higher spatial resolution images, which were partly unavailable and also demand greater processing time and computational costs (Defourny et al., 2019). However, recent advances in satellite technology and their processing capacity, as outlined below, have made high-resolution data more available, encouraging more research on yield estimation methods for the heterogeneous and fragmented landscapes of Europe.

The quantity of available satellite imagery is primarily determined by the satellite's revisit time, i.e., its temporal resolution, which refers to how often the satellite observes the same location. While this is defined by the satellite's orbit, atmospheric conditions such as cloud cover can render optical imagery unusable, as most sensors cannot penetrate clouds. As a result, the interval between two usable observations increases, effectively reducing the actual temporal resolution (Luo et al., 2023). Yet, high temporal resolution is needed for a successful data assimilation. Sisheber et al. (2024) find that a full time series of LAI during the growing season, covering each phenological growth stage, yields the best results compared to using RS data from only a specific growth period. The accuracy of yield estimation declines with a decrease in the number of RS observations (Sisheber et al., 2024).

Since the second half of the 20th century, the number of satellites has rapidly increased. Nowadays, they are divided into two overarching categories: geostationary satellites and polar orbiting satellites (Sabins & Ellis, 2020).

Geostationary satellites are at an orbit height of around 36,000km and provide multiple images per day for one specific area view. This area is always the same and the resolution is low, with one pixel encompassing more than several hundred meters up to kilometers of the ground. Therefore, these images are mainly used for meteorological purposes and are not applicable to model calibration at parcel scale. Instead, polar orbiting satellites orbit the earth at a much closer distance, thereby providing higher resolution, ranging between a couple hundred meters to centimeters, however at a much lower revisit time. Polar orbiting satellites cover the whole surface area of the Earth within a couple of days up to a month, depending on the satellite (Sabins & Ellis, 2020). These satellites are used to understand land interactions, such as vegetation growth and are hence applicable for model calibration.

The earliest satellite mission designed for vegetation remote sensing is the Landsat mission, which since 1972 has sent 9 satellites into orbit that provide images at 30 m resolution in the optical range and 100 m resolution in the thermal range (Sabins & Ellis, 2020). However, with a revisit time of 16 days and the interferences of clouds, the number of usable acquisitions is not dense enough for time series analysis of short agricultural seasons (Defourny et al., 2019).

Alternatively, MODIS (Moderate Resolution Imaging Spectroradiometer) has daily revisit times and creates environmental variables such as the LAI every 8 days, halving the revisit time of Landsat, but at a much coarser spatial resolution of 500m (NASA, n.d.; Stern et al., 2014).

If investigating parcel level yields, this resolution is particularly unsuitable for heterogenous landscapes such as in Europe or African smallholder systems (Defourny et al., 2019).

Therefore, the most suitable and public satellite imagery provider for this study is Sentinel-2 mission of the Copernicus Programme from the European Space Agency. Since 2017, their twin satellites Sentinel-2 A and B orbit the Earth, providing imagery at 10-20 m resolution (depending on the spectral band) with a revisit time of 5 days, considering both satellites. In addition, the Sen4Stat system has been developed to process Sentinel-2 imagery and provide high quality data products, such as the LAI at a resolution of 10m (Deffense et al., 2021). This balance of spatial and temporal resolution, i.e., quality and quantity, makes Sentinel-2-based LAI the most suitable, freely available data source for this study.

2.6 Application examples of process-based models for yield estimation

The models described in section 2.3.2 have been used in different contexts, for different crops, and with slightly differing methodologies to estimate yields. In the following, a selection of previous research is presented to provide an understanding of the performance of the models, by focusing on their application for barley and similar cereals such as wheat.

Upreti et al. (2022) apply the SAFY model for spring barley, winter wheat, and winter oilseed rape at parcel level in Ireland. They used SAFY with a modified soil-water balance module and carbon-flux module. No RS data were used. The crop yield was simulated for three consecutive years, 2011-2013, and evaluated using field measurements by calculating the coefficient of determination R^2, the RMSE and relative RMSE (RRMSE). The grain yield was only calculated for winter wheat and had a low $R^2 = 0.09$ with an RMSE = 2.33 t/ha and RRMSE = 18.6%. In addition, the biomass was evaluated for the winter wheat and spring barley. Winter wheat showed an $R^2 = 0.71$, RMSE = 1.8 t/ha, and RRMSE = 8%, whereas for spring barley, the R^2 = 0.97-0.98, RMSE = 0.7-1.2 t/ha, and RRMSE = 8.6-13.9%.

However, the biomass or DAM (dry aerial biomass) alone is not a good indicator for the actual yield, as shown by Ma et al. (2022): while the modelled DAM has a high accuracy, the modelled yield does not achieve as high R^2 values. Ma et al. (2022) use the SAFY model for plots of winter wheat in the Beijing region, China. At first, a number of parameters are calibrated using in-field DAM and LAI measurements. Then, using LAI values derived from Sentinel-2 three additional parameters are calibrated. Finally, the crop growth is modelled. The correlation between simulated and measured DAM in 2014 (2015), is high with $R^2 = 0.83$ (0.85), RMSE = 1.13 t/ha (1.22 t/ha), and RRMSE = 21.7% (23.7%). However, the yield shows lower accuracies. Comparing the simulated to observed yield, the following results are achieved: R^2 = 0.49 (0.61), RMSE = 1.14 t/ha (1.39 t/ha), and RRMSE = 21.9% (23.3%).

Further application of SAFY has been conducted by Chahbi et al. (2014), in the semi-arid region of Tunisia for wheat and barley. For the period 2010-2012, SAFY was calibrated using in-situ LAI measurements at parcel level, resulting in an $R^2 = 0.35$ for barley, and $R^2 = 0.33$ for wheat between measured and simulated grain yield (Chahbi et al., 2014). Reasons for this low performance could be that crop specific parameters for biomass allocation within the plant (Pla, Plb, and Rs) have not been calibrated for each crop separately. Furthermore, the assumption of SAFY that the productivity of LUE remains constant throughout the growth period could lead to poor yield estimates (Chahbi et al., 2014). However, while the coherence with ground measurements was low, the model was able to simulate the yield dynamics quite well. In the same study, a statistical regression analysis was performed, relating NDVI derived from SPOT5 images to yield, using an exponential equation. This resulted in much better predictions, with R^2 = 0.67 and RMSE = 0.9 t/ha for wheat, and R^2 = 0.62 and RMSE = 0.65 t/ha for barley (Chahbi et al., 2014).

Considering the finding by Chahbi et al. (2014) that a constant LUE is not sufficient in representing crop growth development, Battude et al. (2016) include a mechanism to account for variation in LUE across the season. This mechanism is included in the versions SAFY-CO2, and SAFYE-CO2. Battude et al. (2016) use the adjusted SAFY model in the south of France on maize fields, using Sentinel-2-like LAI for the assimilation for 2013 and 2014. Sentinel-2-like data is created by combining data from the satellites Formosat-2, SPOT4-Take5, Landsat-8, and Deimos-1. They observe a high performance in yield estimation at both the local scale, with R^2 = 0.81, RMSE = 0.93 t/ha, and RRMSE = 8.8, and regional scale, with R^2 = 0.96, RMSE = 0.36 t/ha, and RRMSE = 4.6% (Battude et al., 2016).

The SAFY-CO2 model was applied to winter wheat at two experimental sites in the southwest of France by Pique et al. (2020a), again incorporating remotely sensed LAI using Formosat-2 and SPOT-2/4/5 data for the calibration of certain crop and location specific parameters using a parameter optimization approach. They achieved an R^2 = 0.78, RMSE = 1.02 t/ha, and RRMSE = 21.5%. Furthermore, the model has been applied to corn, soybean, and sunflower in France, Mexico, and Morocco, however, as these are quite different to barley, they are not discussed here.

However, the only implementation of SAFYE-CO2 published at the point of the writing of this study, was on sunflower seeds. Pique et al. (2020b) applied it to sunflower fields in the southwest of France, calibrating it using remotely sensed data. The data was, depending on the year, derived from SPOT4-Take5, Formosat-2, DEIMOS-1, or Sentinel-2a. While the performance for DAM was similar to previous research with RMSE = 0.73-1.18 t/ha and R^2 = 0.74-0.81, the performance for yield was relatively low when considering the R^2 = 0.23-0.31, however, the RMSE = 0.51-0.58 t/ha. Pique et al. (2020b) attribute this finding to the characteristics of the sunflower, such as the grain allocation process, as well as the significant grain loss at harvest.

Next to satellite-derived RS data, process-based models can also be calibrated and updated using RS data from UAVs (unmanned aerial vehicles). These have the advantage of higher spatial resolution and are not disturbed by clouds and atmospheric effects. The spatial resolution contributes in particular to better identification of within-field yield variations (Song et al., 2020). However, UAV data is not as readily available as satellite data. In the following, two examples using RS data from UAVs for yield estimations with the SAFY model are given as a comparative example, as well as to highlight the performance of SAFY.

Song et al. (2020) calibrated the SAFY model based on in-field DAM and LAI measurements of winter wheat fields in Melbourne, Canada. Using simulated observation point clouds from the UAVs, LAI values were generated and used for the final DAM and yield estimation. This allowed them to generate yield maps of submeter to meter scale. While the range of simulated yield is similar to the observed yield, it underestimates yield at the lower and higher ends. Overall, SAFY achieved an RMSE = 0.88t/ha and an RRMSE of 15.22%.

Peng et al. (2021) applied the SAFY model in Inner Mongolia, China at the farm scale, i.e., 75 plots of $9m^2$, where maize was grown. To test SAFY's performance for different water treatments, a soil water content module was added. The calibration was performed with yield observations measured the previous 10 years. UAV-based LAI values were used to update the simulated LAI using sequential assimilation. The simulated yield was mapped at a 2m resolution. Aggregated at the farm scale, the simulated and measured yield were correlated with $R^2 = 0.855$ and an RMSE = 0.7 t/ha. Based on their study, Peng et al. (2021) point out the importance of the quality of the LAI used for the assimilation. Due to its simplicity, SAFY cannot correct for poor LAI values. Therefore, the quality of the LAI will affect the model accuracy, especially during the later stages of crop development where LAI can saturate, meaning the actual LAI is higher than estimated by the RS product, which in turn can lower the simulated productivity. To ensure high quality LAI products, they recommend the use of non-linear LAI inversions.

To compare SAFY with other, more complex process-based models, the following will outline a selection of studies using different process-based models.

Rötter et al. (2012) conducted a model comparison of 9 models: APES-ACE, CROPSYST, DAISY, DSSAT-CERES, FASSET, HERMES, MONICA, STICS, WOFOST for 44 growing seasons of spring barley, at parcel level across different sites in central and northern Europe. The modelers were provided with minimal information for model calibration. The models were evaluated using different metrices, such as the R^2, RMSE, and the Index of agreement (IA). The IA provides a more general assessment of the modeling efficiency, with a range from 0 to 1, where values closer to 1 indicate a higher simulation quality (Rötter et al., 2012). The highest R^2 between simulated and measured grain yield was achieved by the most complex model, i.e., DAISY, with $R^2 = 0.48$. The mean R^2 of all 9 models was below 0.15, i.e., very low. Best in terms of RMSE were HERMES, MONICA, WOFOST with 1.12, 1.28, 1.32 t/ha, respectively. WOFOST, DAISY, and HERMES scored highest in the IA with 0.632, 0.631 and 0.585, respectively. In general, FASSET and CROPSYST performed lowest. Overall, Rötter et al. (2012) conclude that given high uncertainty ranges and inaccurate yield estimates, it is recommendable, if not required to calibrate process-based models according to the region and crop cultivar.

In contrast to the previous study where no RS data was used, Dhillon et al. (2020) use a fusion of Landsat 8 and MODIS data to provide time series of important input variables to ensure a good performance of the tested models, i.e., WOFOST, CERES-Wheat, AquaCrop, CropSYST, and a semi-empirical LUE approach. The study was conducted at the Durable Environmental Multidisciplinary Monitoring Information Network test site in North-East Germany at parcel level for winter wheat biomass. They find that simpler models, which require fewer input parameters, such as AquaCrop and the LUE-approach perform best with an R^2 = 0.86 and 0.83, and RMSE 5.2-5.8 t/ha. In contrast, CERES, CropSYST, and WOFOST, which require a higher number of input parameters, perform slightly worse: R^2 = 0.78, 0.78, 0.77 and RMSE = 5.1-7 t/ha. This difference in performance could be caused by the unavailability of required parameters for the more complex models, highlighting the trade-off of more complex models: while they theoretically could achieve better results, if the input parameters are not available or not well-enough calibrated, then the real performance is not necessarily better.

Silvestro et al. (2017) compared the performance of SAFY to AquaCrop when assimilation of RS data is performed. The study was conducted in a semi-arid region in Central China for winter wheat yield. Yield was calculated at the parcel and district scale. The RS data stems from the Huan Jing Satellite HJ1A/B and the Landsat 8 Operational Land Imager. The RS data was assimilated into the models using sequential assimilation. At parcel level both models performed generally well, yet SAFY showed slightly better performances as well as faster run times. At parcel level the simulated vs measured yield for SAFY achieved an R^2 = 0.53, RMSE = 1.09t/ha, and RRMSE = 18%. AquaCrop had a significantly lower R^2 = 0.34, and RMSE = 1.12t/ha, RRMSE = 24%. At district level, SAFY underestimated the total wheat production by 9.4% (2013) and 12.2% (2014), whereas AquaCrop was highly accurate. A noticeable finding is, that SAFY performs better for low yield values, yet with yields above 6t/ha accuracy decreases, due to underestimations. AquaCrop showed to overestimate low yields, i.e., below 4t/ha, yet also underestimates yields over 6t/ha.

Wallor et al. (2018) emphasize the need of high-quality datasets, in particular for agro-ecosystem models, such as APSIM. By conducting a stepwise calibration of different agro-ecosystem models they show that for most models, among them APSIM, yield estimates improve with each calibration step. The study was performed on two study sites, and APSIM showed strong between-field variability. For one test site, the yield was overestimated and a yield range larger than observed was simulated, but the inter-annual dynamics were well represented. The correlation coefficient averaged across the simulated years for this test site was r = 0.25. For the second test site, the simulated range of yield estimates was lower than observed and the inter-annual dynamics were not well captured. Yet, the correlation coefficient for this site averaged for all the test years was r = 0.46. Overall, this study highlights that site specific high quality data is very necessary to achieve good results with complex agro-ecosystem models.

To sum up, the analysis of the different applications of process-based models, with a focus on the SAFY model versions, has shown the following. The different implementations of the SAFY model versions, have an average performance of R^2 = 0.7 (range = 0.35-0.96), RMSE = 1 t/ha (range = 0.36-1.39 t/ha), and RRMSE = 18 % (range = 4.6-23.3 %). They perform as well, sometimes better than other process-based models presented. Using more complex models does not necessarily improve yield estimation accuracy, due to the difficulty of estimating all parameters accurately (Dhillon et al., 2020) and of adjusting them to local conditions (Wallor et al., 2018). This underpins the advantage of using a less complex model such as SAFYE-CO2 to achieve good yield estimations. As the model version SAFYE-CO2 has few published implementations and none of them occurred in Spain, this study contributes to evaluating its performance as well as transferability to other regions.

One issue that was highlighted is SAFY(E-CO2)'s ability to accurately derive the yield from the DAM. Within the model formulations, the yield is a fraction of the DAM, therefore, processes occurring during senescence are not sufficiently accounted for (Ma et al., 2022; Pique et al., 2020b). This might be a reason for the underestimation of yield observed in several studies (Silvestro et al., 2017; Song et al., 2020).

Concerning the data assimilation, parameter optimization using LAI data is the primary method across the studies and is deemed necessary for good model performance by Chahbi et al. (2014). Rötter et al. (2012) further point out that specifically for large-scale applications, i.e., across different countries, regional and crop-specific calibration is necessary. Furthermore, when using satellite-derived LAI, Peng et al. (2021) underline the importance of deriving it using non-linear methods, as otherwise higher LAI values can be affected by the saturation effect. This finding underpins the suitability of the Sen4Stat tool, as it implements an Artificial Neural Network trained using the inversion of a Radiative Transfer Model (Deffense et al., 2021), making it a non-linear method.

While using UAVs for image acquisition might be useful for detecting within field variability, it is not a feasible approach for larger scale applications, such as at country level. Therefore, remotely sensed data is used in most studies. As most of the studies were conducted before the launch of the Sentinel-2 satellites, they used data from other satellites of different spatial resolutions - some higher, some lower - than Sentinel-2.

However, one issue not yet addressed is the model's operationalization. Most of the previously mentioned studies were conducted on experimental sites or few fields. But if models like SAFYE-CO2 are supposed to become operational, that is, to be used for national, regional, or global yield estimation, significantly more computing power will be required (Luo et al., 2023). Furthermore, yields are often expected to be provided pre-harvest to allow for better planning. These two challenges of spatial and temporal scalability for operationalization of crop growth models are addressed in the following section.

2.6 Operationalization and scalability of process-based models

2.6.1 Temporal availability of yield estimation

Crop yield can be estimated or forecasted. Yield estimation refers to calculating yield at the end of the harvest, whereas yield forecasting predicts the final yield before the field is harvested (Basso and Liu, 2019).

The previous examples involved yield estimation after the harvest, as accuracy is higher when data from the entire season can be used (Sisheber et al., 2024). Weather conditions, such as excessive rainfall, can still affect yield during the final development stages (van der Velde et al., 2012). While calculating yield at the end of the season is sufficient for purposes such as identifying long-term trends or for research purposes, farmers, governments, traders, and other stakeholders are more interested in yield forecasting (Basso & Liu, 2019; van der Velde et al., 2019).

The advantage of yield forecasts is that they allow farmers to adjust farming practices, and enable governments and traders to set prices and regulate imports and exports in advance to avoid food shortages (Weiss et al., 2020). Van der Velde et al. (2019) analyzed the use of the European Commission's Monitoring Agriculture with Remote Sensing (MARS) bulletin, which is published on a monthly basis. They found that demand for the bulletin peaks June, i.e., one month before the harvest of most winter cereals.

Regarding the feasibility of pre-harvest yield estimation, Sisheber et al. (2024) investigated the effectiveness of calibrating SAFY when using LAI data from different growing stages. They found positive results, showing that calibrating SAFY with LAI data until the peak LAI is sufficient to achieve accurate yield estimates. Using the full LAI curve resulted in an RMSE = 1.05 t/ha, RRMSE = 14.8%, R^2 = 0.77. Using the LAI curve only until peak LAI achieved RMSE = 1.17 t/ha, RRMSE = 16.4%, R^2 = 0.68. In contrast, using the post-peak LAI curve, the prediction accuracy significantly decreased, with RMSE = 2.14 t/ha, RRMSE = 30%, and R^2 = 0.35. This finding is likely because the modeled yield is primarily determined by the duration of green biomass formation, i.e., green leaves, which takes place from heading until the grain-filling stage, corresponding to the peak LAI (Huang et al., 2015, Tewes et al., 2020b as cited in Sisheber et al., 2024). This finding is reproduced by Chahbi Bellakanji et al. (2018) who used the SAFY model to relate a 10-day period of maximum LAI to yield using a linear regression, resulting in R^2 = 0.77 and RMSE = 537 kg/ha.

These results suggest that using the SAFY (and SAFYE-CO2) model can provide yield forecasts at the most critical time, i.e., one month pre-harvest, without substantially compromising accuracy.

2.6.2 Spatial availability of yield estimates

In addition to national statistic offices which primarily use sampling survey methods for post-harvest yield estimation, as outlined in Section 2.2, pan-national and even global initiatives have been set up to estimate and forecast yield. Fritz et al. (2019) provide a comparison of these different initiatives that are briefly outlined below, focusing on their main characteristics.

The oldest initiative is the FAO’s Global Information and Early Warning System, established in the 1970s, which provides global estimates for crop production, amongst others. It combines meteorological data, remotely sensed vegetation indices, further auxiliary data (e.g., cropland maps), crop models and statistical models, to provide three annual reports on global food production as well as special alerts and reports on areas where production is critical.

The Famine Early Warning Systems Network led by USAID and implemented since 1985, covers monthly yield production for the world’s 36 most food insecure countries. It mainly relies on anomaly analysis using agro-climatology data.

For European countries, the MARS Crop Yield Forecasting System was developed under the European Commission in 1992. It provides monthly bulletins on yield forecasts which are calculated using a statistical model based on primarily historical yields and cropland area, meteorological data, crop model simulations, and satellite-derived biophysical parameters.

A Chinese counterpart is CropWatch, developed by the Institute of Remote Sensing and Digital Earth at the Chinese Academy of Sciences since 1998. It provides sub-national crop production estimates for the 9 largest countries, national production for 31 countries producing 80% of the world's staple crops, as well as regional and global production forecasts. Depending on the scale, different predictors such as meteorological data and satellite-derived vegetation indices, are used to implement agro-meteorological models and regression analyses (Fritz et al., 2019; Wu et al., 2014).

Lastly, GEOGLAM monitors crop status through remotely sensed data, which is input into the Agricultural Market Information System. This started in 2011 as an international effort for more transparency in the global crop market, covering 29 countries that produce 80-90 % of the world's four staple crops (Becker-Reshef et al., 2019; Fritz et al., 2019).

Hence, multiple information systems and tools are available to monitor global crop yield and production at national scale. However, these estimates aggregated at national level do not come without their limitations. Firstly, spatial differences at sub-national scale are concealed (Paudel et al., 2022), which could hide relevant information. Furthermore, the predictors used for the yield estimation, such as meteorological data or vegetation indices, are aggregated at national scales. This leads to a loss of information, making yield estimations less accurate than when calculating yield at smaller scales, e.g., parcel level, to be then aggregated to national level (Chipanshi et al., 2015).

To improve yield estimates and make them available at multiple scales, from parcel to nation, an alternative would be to apply crop growth models at parcel or pixel scale. However, crop growth models are very computationally expensive, and using them for parcel or pixel level yield estimation for a whole country, let alone at the global scale, would be nearly impossible (Li et al., 2024). Furthermore, due to advancements in machine and deep learning algorithms which provide high yield estimation accuracies, Baker et al. (2018) are questioning whether process-based models will be needed in the future due to their complex implementation. However, they conclude that leveraging the complementary strengths of statistical and process-based models should be the way forward. Two such approaches have been developed and are outlined in the following.

In 2015, Lobell et al. (2015) published the Scalable Crop Yield Mapper (SCYM) which uses the APSIM model as a basis for a statistical regression. The idea behind the approach is that instead of using field data to establish a statistical regression, the outputs of multiple APSIM runs are used. First, the APSIM model is run for different weather conditions, years, and different parameterizations to simulate a large range of possible crop growth curves. In a second step, a simulated variable that can be observed from satellites, such as the LAI, and the simulated yield are used to train a regression model. Finally, the regression can be applied pixel-wise to the region of interest, using remotely sensed data and weather data as inputs. Consequently, this method combines the benefits of both statistical and process-based models. It reduces the need for extensive collection of training data, while being able to encompass extreme weather conditions, e.g., by using extreme weather predictions as inputs for the model. And at the same time significantly reduces computing costs that would have been required when applying the crop growth model pixel-wise. Since 2015, the SCYM has been applied to different regions, such as the US, China, but importantly, also to small-holder farming systems such as in Uganda (Deines et al., 2021; Li et al., 2024; Lobell et al., 2020).

A similar yet different approach has been developed by Wijmer et al. (2024). They created AgriCarbonEO (ACEO), a processing chain, at the core of which is the SAFYE-CO2 model. Instead of calculating a regression, the ACEO creates a Look-Up-Table (LUT). For each weather cell (and possibly soil type found within the weather cell), a user-defined number of simulations is run with different parameterizations. Then, pixel-wise remotely sensed LAI data is compared to the LAI simulations in the LUT using Bayesian Normalized Importance Sampling via LUT Generation (BASALT). This computes the likelihood for each of the parameter sets explaining the pixel's LAI curve and provides a final posterior distribution of parameter values per pixel. As a result, not only is one LAI curve with the corresponding yield determined per pixel, but an uncertainty range is also provided.

Depending on the number of LUT entries generated per weather by soil grid cell, as well as the number of fields within each weather by soil grid cell, this approach can significantly reduce the processing power required to use a process-based model at pixel level for a large area, such as province or even country. The main processing cost is required for the generation of the LUT. Afterwards, each pixel's yield and LAI determination process is fast.

Since the SAFYE-CO2 model does not only calculate yield, but also the water and carbon balance, scaling the model across large regions appears promising for diverse agricultural and environmental applications. However, the ACEO processing chain is a very new approach, and has so far only been applied to estimate the carbon exchange on wheat fields in the southwest of France (Wijmer et al., 2024).

To conclude, the literature review has shown that the SAFYE-CO2 model strikes a good balance between simplicity and biological realism, with performances similar to other models, can be operated at large scale and in forecasting mode, and due to its process-based nature, is likely suited to account for future meteorological changes.

Therefore, this study aims to apply the SAFYE-CO2 model to the province Castilla y León in Spain to estimate and forecast barley yield at parcel level, as well as to investigate the performance of the ACEO at pixel level to assess the scalability of the model.

3. Objectives

The overarching goal of this study is to assess and compare the performance of different methodologies for estimating barley yield at parcel level in the Spanish province of Castilla y León. A focus is put on the implementation of the process-based model SAFYE-CO2. In addition, the study investigates the model's spatial scalability while preserving high-resolution outputs using the ACEO processing chain, as well as its potential to be used for yield forecasting. Furthermore, SAFYE-CO2's performance is compared to the performance of the statistical Sen4Stat model. More specifically, the objectives are as follows:

A) Performance assessment of the SAFYE-CO2 model at parcel level:

1. Assessment of potential uncertainty caused by model parameterization through a sensitivity analysis, and design of a process to calibrate parcel-based parameters using satellite-derived LAI assimilation.
2. Assessment of SAFYE-CO2's performance in combination with satellite-derived LAI assimilation for yield estimation at parcel level.
3. Assessment of SAFYE-CO2's performance in combination with satellite-derived LAI assimilation for yield estimation aggregated at provincial level.
4. Analysis of SAFYE-CO2's potential for yield forecasting by restricting LAI assimilation up to one to two months pre-harvest.

B) Performance assessment of the ACEO processing chain's yield estimation at pixel level, aggregated to parcel level.

C) Application of the statistical model Sen4Stat for yield estimation at parcel level.

4. Methodology

4.1 Study area

Spain is Europe's third largest producer of barley with one third of its production being located in Castilla y León. Castilla y León is an autonomous community in Spain, located between 40° and 43°N latitude and 1° to 7°W longitude, situated on the Meseta Central, the interior plateau of the Iberian Peninsula (see Figure 2). The region experiences a continental Mediterranean climate characterized by wet, cold winters and dry, warm summers (Climate Change Knowledge Portal, n.d.). This for Spain comparatively cold and dry climate, typical of the northern Meseta Central, combined with shallow, loamy soils (Ballabio et al., 2016), provides favorable conditions for barley cultivation (Martínez-Moreno et al., 2024).

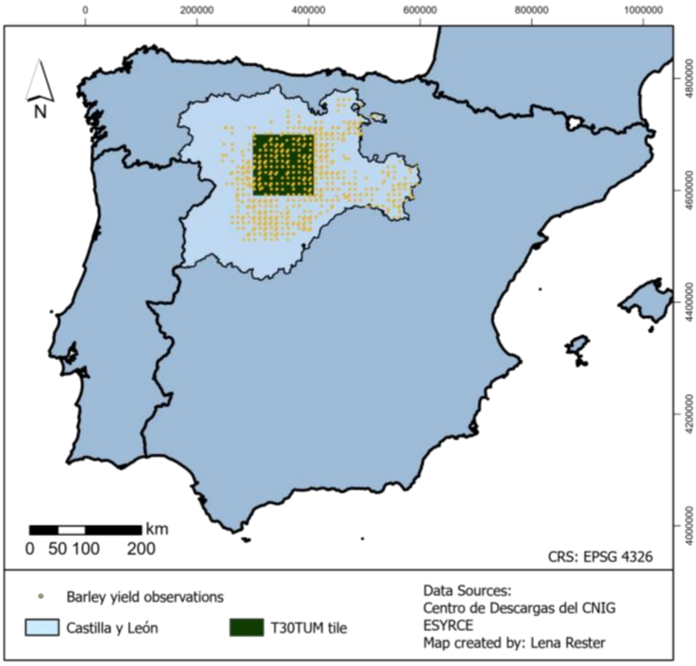


Figure 2: Study area Castilla y León, displaying the barley yield observations from ESYRCE and the extend of the T30TUM tile used for the ACEO implementation.

The 2020 barley growing season was characterized by above-average rainfall and temperatures during key growing periods, without the occurrence of heatwaves that could damage crops. Hence, the 2020 harvest outperformed the five-year yield average across much of Spain, including CyL (Baruth et al., 2020).

4.2 Materials

4.2.1 Yield Data

Yield data were obtained from the "Encuesta sobre Superficies y Rendimientos de Cultivos" (ESYRCE), the annual Survey on Crop Surfaces and Yields conducted by the Spanish Ministry of Agriculture, Fisheries and Food (Ministerio de Agricultura Pesca y Alimentación, 2020). Data collection is carried out using a systematic sampling design. For every 100 km² block - corresponding to a 10x10 km UTM grid - three segments of approximately 700 m x 700 m each are sampled (Ministerio de Agricultura Pesca y Alimentación, 2020). Within each segment, the land is divided into individual parcels (see Figure 3). Field surveys are conducted by experts in May and August to collect information on cultivated area, yield estimates, and crop variety, among other variables (Ministerio de Agricultura Pesca y Alimentación, 2020). The yield estimation method can be one of the following methods: (1) visual (eye) estimation, (2) farmer-reported data, (3) combine harvester measurements, (4) ear and grain counts, (5) counting and weighing of ears, (6) weighing of ears, (7) counting and weighing of fruits or tree counts, and (8) other sector-specific methods (Ministerio de Agricultura Pesca y Alimentación, 2020).

In this study, the results of these field surveys are used to identify barley-producing parcels and to provide calibration and validation data for the yield estimation process. In 2020, yield data were available for 2,552 winter barley parcels. As in CyL two-row barley accounts for 94% of the total barley-producing area, it is reasonable to assume that parcels are likely cultivated with this variety (Ministerio de Agricultura Pesca y Alimentación, 2020). Additionally, the majority of parcels were rainfed, with a small subset being irrigated.

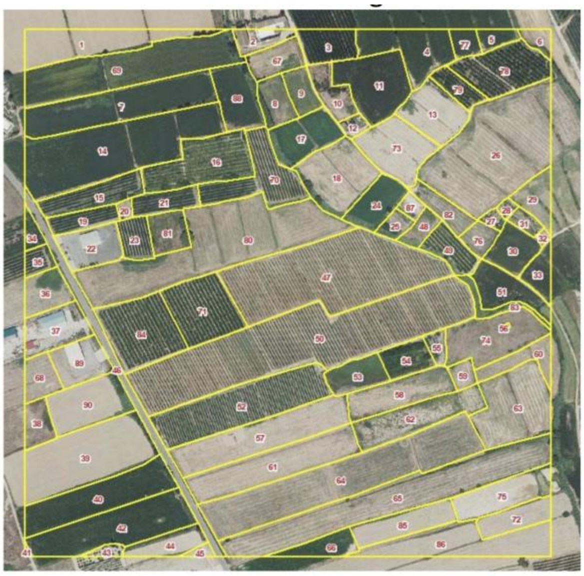


Figure 3: Example of a 700 m x 700 m segment with individual parcels to be sampled (Ministerio de Agricultura Pesca y Alimentación, 2020)

4.2.2 Leaf Area Index Data

As already explained in Section 2.4, the LAI represents the amount of green leaf area available for photosynthesis and other exchange processes between the canopy and the atmosphere (Deffense et al., 2021). Its strong linkage to plant productivity, the model's biomass-related parameters, as well as its availability from remote sensing observations, make LAI a valuable variable for model calibration through data assimilation.

As the satellite only measures the green leaves, it is often also referred to as the Green Area Index or Green Leaf Area (GLA), the latter being the term used for the corresponding model variable. In this study, GLA and LAI are used synonymously with GLA mainly referring to the model output and LAI to the satellite-derived values.

The observational LAI data were retrieved from the Sen4Stat toolbox, a processing chain that utilizes Sentinel 2 Level 1C and 2A (and optionally Landsat 8 Level 1T) data (Bontemps et al.,

2024) as inputs to BV-net, a trained artificial neural network, to derive biophysical variables such as LAI (Deffense et al., 2021). The final product is a CSV table which contains for each parcel the mean LAI, the standard deviation, the number of valid and invalid pixels for each satellite acquisition. The data processing involved LAI value extraction, outlier removal, temporal gap filling, and smoothing, as illustrated in Figure 4 and outlined below.

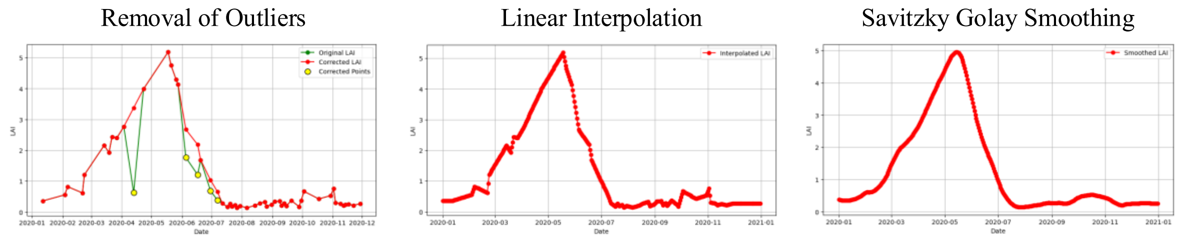


Figure 4: The processing steps of LAI cleaning, i.e., removal of outliers, linear interpolation and smoothing, for one parcel.

For each parcel the mean LAI was used only if the number of valid pixels was at least 80% of the total pixels. Despite this threshold, there were obvious outliers in the form of sudden drops, which especially at high LAI values are unlikely to reflect actual vegetation changes and are often attributed to factors such as clouds, atmosphere, measurement errors, geometric misregistration, artifacts due to data resampling, or anisotropic reflectance effects (Cai et al., 2017). However, traditional outlier detection techniques such as Z-score, standard deviation, or Local Outlier Factor are inadequate in this context, as they often flag high but valid LAI values rather than the sudden, contextually implausible drops. Misapplication of these methods could lead to the removal of peak LAI values, severely underestimating yield during subsequent analysis.

Instead, a contextual approach was used to detect and correct outliers. The algorithm compared each LAI value to the preceding one. If the LAI exceeded 1 and dropped by more than 25% between two consecutive dates, the latter value was flagged as an outlier. It was then replaced with the average of its two neighboring values. The algorithm then continued to the next value in sequence, applying the same logic. The threshold of LAI > 1 was applied to avoid falsely correcting natural variations at low LAI values, which are later addressed through smoothing.

After the outlier correction, a linear regression is applied to fill gaps between the observational dates. During the growing season, successful LAI observations occurred only 3–4 times per month. However, for effective model assimilation, a higher temporal resolution is desirable. On managed agricultural fields, the LAI curve is expected to follow a predictable phenological pattern and therefore a simple linear interpolation can be applied to generate an LAI value for each date.

As a last step, a smoothing function is applied to reduce residual noise. Various smoothing methods exist, and there is no one best method, as stated by Cai et al. (2017). While all methods reduce noise and improve signal quality, they are suitable for different use cases. Global methods, such as spline smoothing or asymmetric Gaussian function fitting, fit a function to the provided dataset (Cai et al., 2017). Therefore, they are useful for reconstructing more generalized, seasonal curves or in cases of few or low data quality, but not for capturing seasonal details (Cai et al., 2017). Local methods, such as Savitzky-Golay or LOESS (locally estimated scatterplot smoothing), on the other hand, should be used when aiming at capturing intra-seasonal variations, as they adapt the data stepwise and locally (Cai et al., 2017). Thereby, local details and short-term variations are better captured. As this study focuses on one season of one crop type, preserving temporal detail and curve shape was critical. Therefore, Savitzky-Golay smoothing was selected, a commonly applied and straightforward method which smooths the data while retaining key signal features (Savitzky & Golay, 1964). For each point, a polynomial is fitted to a moving window of neighboring values using a least-squares approach. The center point of this fit replaces the original value (Savitzky & Golay, 1964). Two user-defined parameters influence the outcome: the polynomial degree and the window size. Lower-degree polynomials (e.g., 1 or 2) result in smoother curves, whereas higher degrees (e.g., 5 or above) retain more variability but may overfit and preserve noise. In this study, a polynomial degree of 3 was used to strike a balance between noise reduction and detail preservation. The effect of the window size is illustrated in Figure 5. Due to the initial few data points (3-4 per month), a window size of 41 was chosen to cover at least one month of data points.

These steps yielded, for each parcel, a temporally continuous and noise-reduced LAI curve, with one smoothed LAI value per date.

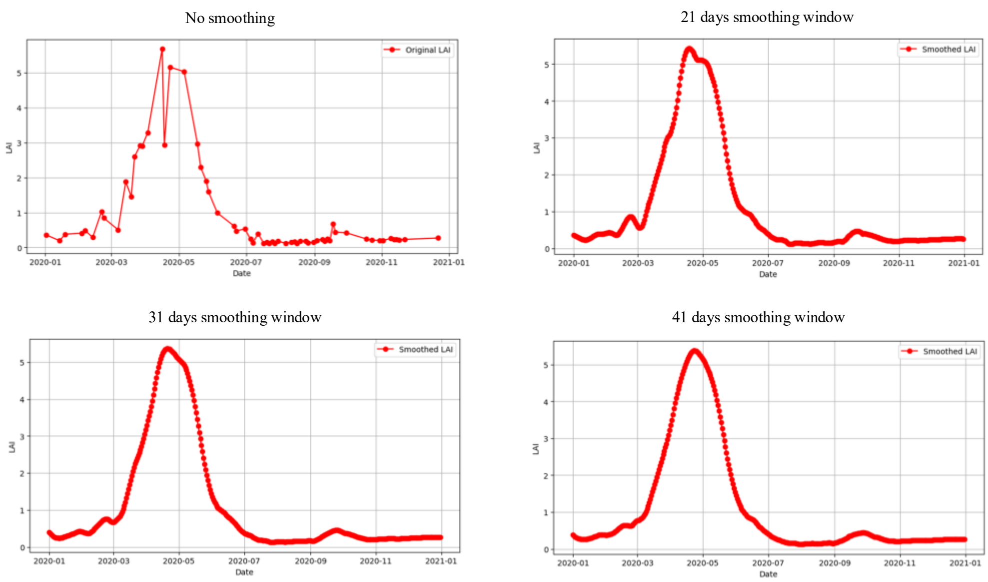


Figure 5: Effect of different window-sizes on the degree of smoothing using for Savitzky-Golay smoothing.

4.2.3 Data Cleaning

Data cleaning was performed based on the crop quality rating and parcels' maximum LAI values. Parcels with a crop quality rating of 2, e.g., due to implausibly low yields of 1 kg/ha, were excluded as outliers due to their unrealistic values. In addition, parcels were investigated based on their maximum LAI during the season. As displayed in Figure 6, the maximum LAI values ranged from less than 0.5 up to 8 with the majority of parcels having a maximum LAI between 1 and 4. While it is possible for barley fields, in particular in the Mediterranean region, to have low LAI values (Berjón et al., 2003; Garrachón-Gómez et al., 2024), a maximum LAI value below 1 is highly unlikely. Therefore, parcels with such low maximum LAI values were investigated using the LAI time series and Sentinel-2 imagery. This showed that these parcels did not have the expected LAI curve which was confirmed in the Sentinel-2 NIR/SWIR/Red false color composites, as displayed in Figure 7. Therefore, parcels with an LAI < 1 were removed.

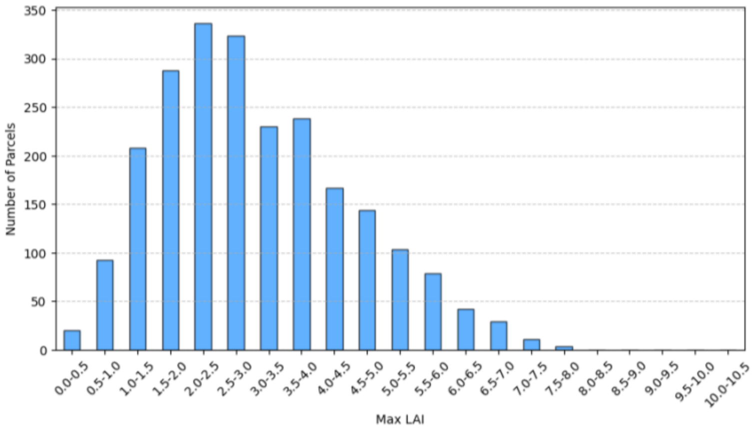
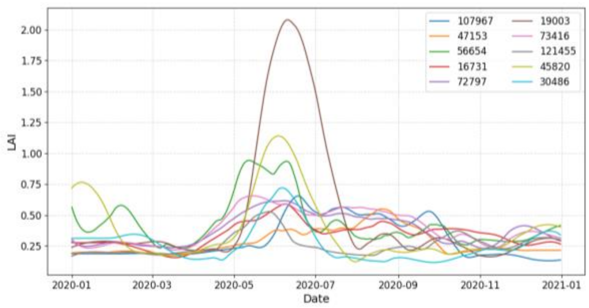


Figure 6: Distribution of the seasonal maximum LAI values across all barley parcels.

A) 

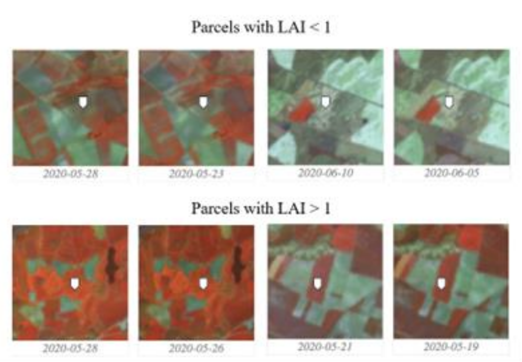
B) 

Figure 7: A) Comparison of LAI curves of parcels with a maximum LAI < 1 to a parcel with maximum LAI > 1. B) Comparison of 2 parcels with high and 2 parcels with low maximum LAI during the peak growing season.

Due to method-specific biases (see Literature Review, Section 2.2), as well as different growth patterns under irrigated conditions, the parcels were divided into three groups, as shown in Figure 8. This approach aims to maintain uniformity in data quality, given the varying levels of precision associated with different estimation methods and to evaluate the model‘s ability to simulate irrigated parcels. Moreover, the largest group, i.e., eye-estimated rainfed parcels, were divided into two groups based on their harvest time, i.e., June or July. These groups were used to assess the ability of the model to work with a restricted LAI dataset. For each group, a set of calibration and validation parcels were selected for the calibration of the Harvest Index.

Figure 9 shows the yield distribution for all parcels and calibration subsets, as well as parcel sizes by group.

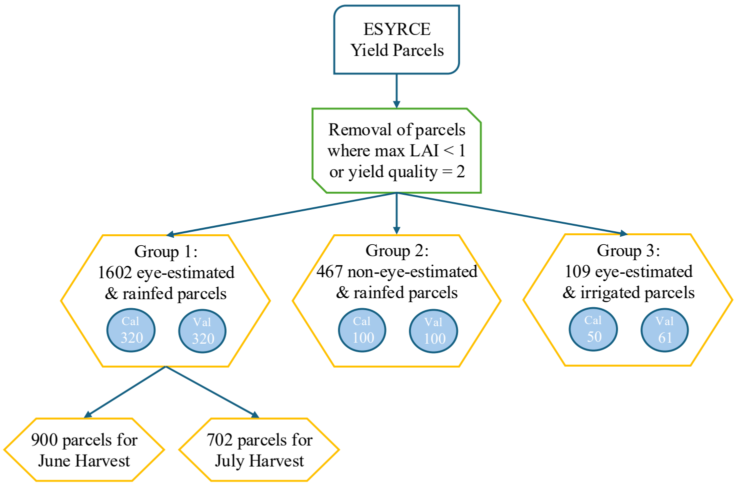


Figure 8: Division of yield data based on estimation method and irrigation.

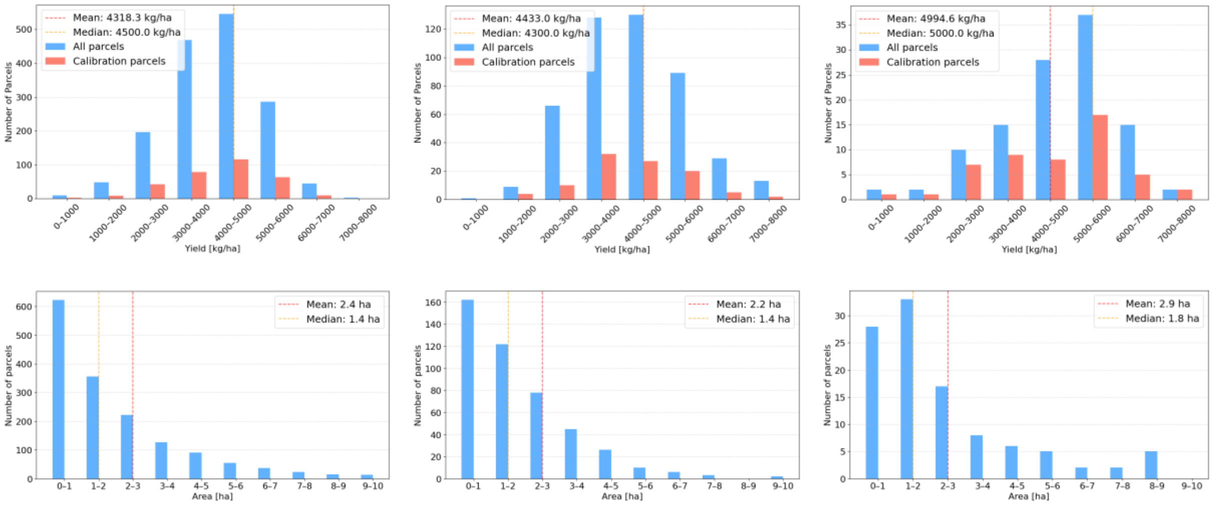


Figure 9: Distribution of yield and parcel size for each parcel group (top: yield in kg/ha, bottom: parcel size in ha; left to right: group 1, group 2, group 3).

4.2.4 Data on Soil Hydraulic Properties

To improve the accuracy of the plant's water availability estimates, three soil parameters used by the model were initialized using corresponding datasets.

For the soil moisture parameter, the dataset "Surface Soil Moisture 2014-present (raster 1 km), Europe, daily – version 1" was obtained from the Copernicus Land Monitoring Service (European Union's Copernicus Land Monitoring Service information, 2025). It provides the relative water content of the upper soil layer, expressed as percent saturation (Bauer-Marschallinger & Paulik, 2019). It covers continental Europe at a spatial resolution of approximately 1 km, with a revisit frequency of 1.5 to 4 days depending on location (Bauer-Marschallinger & Paulik, 2019). Since the data is used only to initialize the model, values from the start of the simulation, i.e., the 1st of January were used.

As the model uses absolute, not relative, water content, the following conversion using the soil porosity, as suggested by Bauer-Marschallinger & Paulik (2019), was applied:

$$SM_i = \frac{WC_r * p}{100} \quad 1$$

With i indicating the soil layer, where Layer 1 = 0-10 cm, Layer 2 = 10-35 cm, Layer 3 = 35-150 cm, and with p calculated according to Weil & Brady (2016) as:

$$p = 1 - \frac{p_b}{p_p} \quad 2$$

With p_b being bulk density downloaded from the European Soil Data Centre (ESDAC) (Ballabio et al., 2016; Panagos et al., 2022) and p_p being the particle density, which is estimated to be 2.65 g/cm3 for mineral soils (Weil & Brady, 2016).

Moreover, the model requires information on the wilting point and field capacity, which are key parameters describing soil water availability for plants. More specifically, the field capacity measures the amount of water retained in the soil against gravitational forces, typically measured two to three days after wetting the soil profile (Veihmeyer & Hendrickson, 1931). The wilting point indicates "the soil moisture content at –15848 cm matric potential" (Tóth et al., 2015) beyond which plants can no longer extract water. These datasets were obtained from the ESDAC using the "Maps of indicators of soil hydraulic properties for Europe" dataset from 2016 (Panagos et al., 2022; Tóth et al., 2015). The dataset provides the wilting point and field capacity at a spatial resolution of 1 km across the European continent, derived using the pedo-transfer functions developed by Tóth et al. (2015). No further processing was necessary, as the data were already expressed in the required unit (cm^3/cm^3, i.e., dimensionless).

4.2.5 Meteorological data

As an agro-meteorological model, SAFYE-CO2 requires daily input data for precipitation, mean air temperature, solar radiation, and evaporation. These variables were obtained from the ERA5-Land hourly reanalysis dataset (1950–present), provided by the Copernicus Climate Change Service via the Copernicus Climate Data Store at a spatial resolution of 9 km. The ERA datasets are widely used across the climate change literature as they provide many relevant climate variables from past to present across the globe (Hersbach et al., 2020).

Thus, for the season January to December 2020, data on 2 m air temperature, total evaporation, surface downward shortwave radiation, and total precipitation was downloaded for the study area, within the geographical extent of 40°N to 43.3°N and 1.7°W to 7.1°W. The ERA5-Land hourly data was processed as follows.

ERA5-Land provides data at hourly steps but the model reads in daily values. Therefore, for evaporation and solar radiation, the daily sum of the hourly values was calculated, whereas for the temperature, the daily mean was calculated. For precipitation, ERA5-Land records cumulative hourly totals, with the 00:00 value representing the total for the previous day (ECMWF, 2024). Next, SAFYE-CO2 also requires diffuse radiation, which was calculated using a script included in the SAFYE-CO2 model. It calculates the Top-Of-Atmosphere (TOA) solar radiation based on the day of year (DOY) and latitude. Then, following Bindi et al. (1992) and de Jong (1980), diffuse radiation is calculated based on the amount of direct radiation, which is the ratio of downwelling solar radiation provided by the ERA5-Land data to the calculated TOA radiation. Consequently, when direct radiation is high, diffuse radiation is low, and vice versa.

4.3 The SAFYE-CO2 Model

As outlined in Section 2.3.2, SAFYE-CO2 is the latest version of the SAFY-models. More specifically, the version 2.0.5 is used, which is distributed by the Centre d'Etudes Spatiales de la Biosphère (CESBIO) (Ceschia et al., n.d.). This version calculates daily leaf and grain production (i.e., yield) as well as water and carbon fluxes and stocks. Furthermore, it explicitly accounts for the effects of water and temperature stress on crop productivity (Pique et al., 2020b). While the full model framework and its functions have been described in detail by previous studies (Duchemin et al., 2008; Pique et al., 2020a; Pique et al., 2020b), the following section highlights the processes and components most relevant to this study and the flowchart in Figure 10 provides an overview of the model structure.

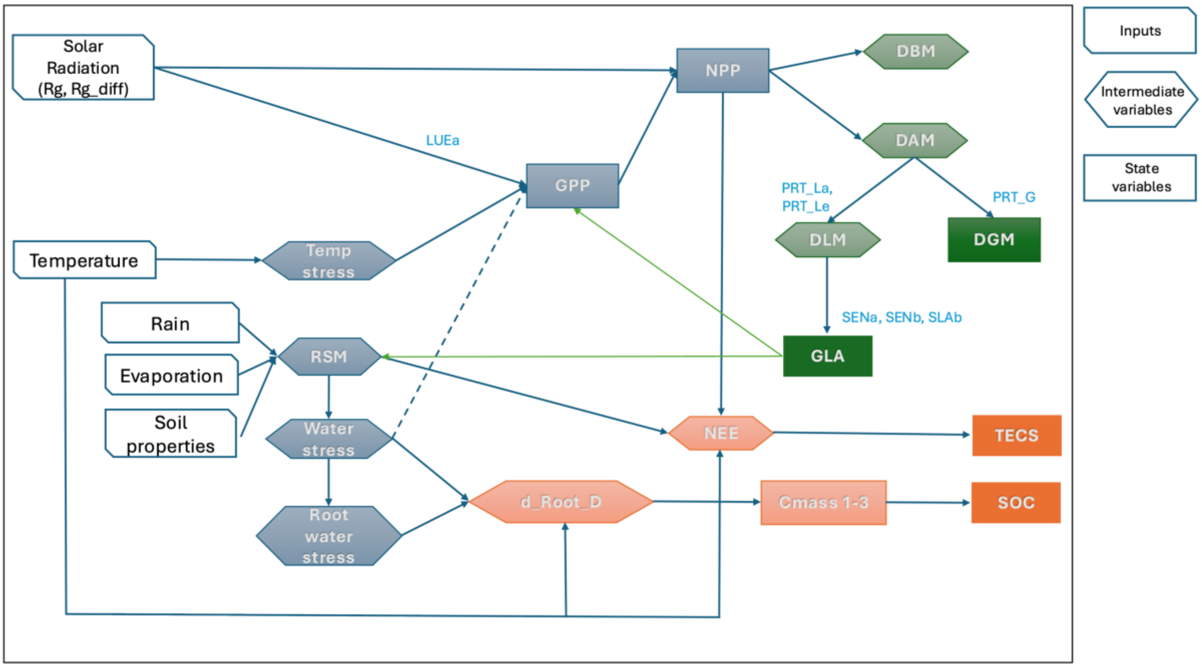


Figure 10: Generalized flowchart of the SAFYE-CO2 model illustrating the inputs, intermediate processes, and state variables. Orange boxes highlight the processes related to carbon stocks and fluxes (TECS = Total Ecosystem Carbon Stock, SOC = Soil Organic Carbon, NEE = Net Ecosystem Exchange, d_Root_D = daily change in root depth), while green boxes represent the production of plant biomass and yield (DGM). Parameters subject to calibration on a field-specific basis (except for PRT_G) are annotated in blue.

At the start of each simulation, the model reads the necessary input files for each parcel. These include (1) a parameter file containing soil and vegetation parameters, some of which are derived from the aforementioned soil datasets and others subject to calibration, (2) a forcing file that provides the required daily weather variables: air temperature, precipitation, evaporation, downwelling solar radiation, and diffuse solar radiation, and (3) a configuration file specifying the start and end dates of the simulation period.

The core of the model is organized into five modules, which are executed sequentially for each day of the simulation period and described in the following sections by highlighting the parameters that have to be calibrated on a parcel-level basis.

4.3.1 Soil Module

The soil module is called first for each timestep and updates the distribution of carbon and water across the three soil layers, accounting for root growth dynamics. These calculations update both the absolute soil water content and the relative soil moisture for each layer which can be used to calculate water stress in the second module. However, based on the recommendation of the model designers, this was excluded. Hence, this module is most relevant when calculating carbon stocks and fluxes.

4.3.2 Plant Physiology Module

This module focuses on estimating biomass production, primarily through the calculation of Gross Primary Productivity (GPP) between the parcel-specific emergence and harvest days:

$$GPP = PAR * LUE * stress * SR10 \tag{3}$$

Where *stress* is a value between 0 and 1, with 1 indicating no stress, and *SR10* is a factor which decreases canopy photosynthesis capacity during senescence. *PAR* is the amount of photosynthetically active radiation that intercepts the plant's canopy, i.e., the amount of radiation used for photosynthesis, calculated using Beer's law:

$$PAR = \; Rg_{PAR} * Rg * (1 - e^{-Kex*GLA}) \tag{4}$$

Where Rg_{par} and *Kex* are a-priori parameters, *Rg* is the amount of incoming solar radiation and *GLA* the existing photosynthetically active Green Leave Area. This equation highlights the feedback loop of GLA as it influences the total biomass production, while being an output of the biomass production.

Light Use Efficiency indicates the efficiency with which PAR is converted into carbon-biomass, calculated as follows:

$$LUE = LUE_a * \; e^{Rg_{diff}*LUE_b} \tag{5}$$

Where LUE_b is an a-priori defined parameter, Rg_{diff} the incoming diffuse radiation, and LUE_a one of the parameters that needs to be calibrated. LUEa is considered as one of the most crucial parameters, as it describes the efficiency with which light is transformed into biomass and often considered as being able to account for temperature, water, and other stress factors.

In the next step, autotrophic respiration is calculated, comprising maintenance respiration, which is required to sustain existing plant tissues, and growth respiration, which accounts for the energy expended to develop new biomass. Subtracting total respiration from GPP yields the Net Primary Production (NPP), representing the net amount of carbon assimilated and retained by the plant. The NPP is the main output of this module, needed for the distribution of biomass across the different organs in the next module.

4.3.3 Plant Phenology Module

As displayed in Figure 10, this module concerns itself with the allocation of the daily produced new biomass, i.e., ΔNPP, to the dry below ground biomass (DBM) and dry above ground biomass (DAM). DAM is further partitioned into dry leaf biomass (DLM) and dry grain biomass (DGM). To assimilate LAI and estimate yield, the following equations are the most relevant. First, the allocation of biomass to leaves is calculated as:

$$DLM = DLM_{t-1} + PRT_L * (\Delta NPP - DBM) \quad 6$$

Where PRT_L is the coefficient defining how much of the aboveground biomass is allocated to leaves versus grain, which is determined by sum of efficient temperature since emergence (*SMT*), and the two parameters *PRT_La,* which influences the proportion of DAM that is not allocated to leaves, and *PRT_Le,* which determines at what point no new leaves are formed. These two parameters have a critical impact on the leaf formation and hence, the GLA curve, therefore they are calibrated parcel-specific.

The DLM is then transformed into GLA, considering the growth and senescence of leaves:

$$GLA = GLA_{t-1} + GLA_{grow} - GLA_{sen} \quad 7$$

$$GLA_{grow} = \Delta DLM * SLA_b \quad 8$$

$$GLA_{sen} = GLA * \frac{SMT - SEN_a}{SEN_b} \quad 9$$

Where *SLAb* is a parameter used to convert the dry leaf biomass into the available green leaf area based on the Specific Leaf Area (SLA) property of the plant, while *SENa* indicates the SMT at which senescence starts, and *SENb* controls the rate of leaf senescence. The three parameters have to be calibrated per parcel, in particular since SENa and SENb or the main parameter controlling the senescence curve of the GLA.

Lastly, this module also allocates the DAM to grain biomass (DGM, i.e., yield) using the equation:

$$DGM = DGM_{t-1} + PRT_G * (\Delta NPP - DBM) \quad 10$$

The only parameter specifically influencing DGM allocation is PRT_G, which in this study is interpreted as the harvest index (HI), as it defines the proportion of daily increase in NPP allocated to grain. This is a simplistic yet commonly method to relate biomass to the yield, but it cannot account for stress factors during the senescence period that impact grain production. The PRT_G is calibrated for all parcels universally.

4.3.4 Heterotrophs and Farming Practices Module

The heterotrophs module estimates heterotrophic respiration, making it a key component in tracking soil carbon dynamics. The farming practice module allows the incorporation of carbon imports, such as fertilizer, however since there is no information on carbon imports and the goal of this study is not to optimize carbon modeling, this function is not used. Furthermore, irrigation practices can be included, yet, since information on irrigation practices on the few parcels that are being irrigated was not available, this function is not included either.

4.3.5 Model Application

After a model run, daily values for variables such as GLA, DGM, GPP, or NPP are provided. The main variables used in this study are the GLA, and DGM, which indicate the leaf area observable using remotely-sensed LAI and the amount of possible yield.

An overview of the model application process, from parcel-specific parameter calibration using the LAI assimilation to HI calibration is provided in Figure 11. The final output constitutes yield values per parcel. In addition to parcel-level analysis, yield values are analyzed at the provincial level by grouping the parcels per province and calculating the average yield.

In the following, the assimilation of remotely sensed LAI to calibrate the parcel-specific parameters as well as the calibration of the HI are outlined in more detail.

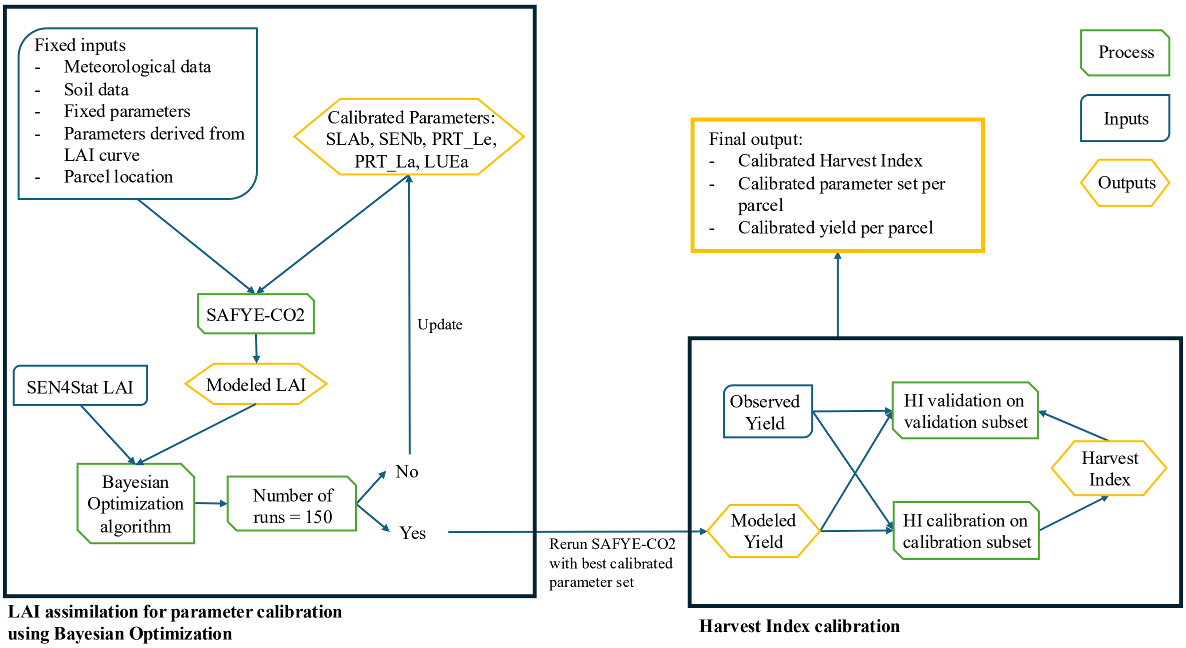


Figure 11: Representation of the full modeling process, from LAI assimilation to calibration of the Harvest Index.

4.4 Remote Sensing assimilation for parcel-specific parameter calibration

4.4.1 Identification of parameter values and ranges

The choice of parameters to be calibrated was based on previous work by Pique et al. (2020a) and Pique et al. (2020b), who used LAI assimilation to calibrate the parameters LUEa, emerg (day of emergence), SLAa (equal to SLAb in this model), PRT_La, PRT_Lb (equal to PRT_Le), SENa, and SENb on a per-parcel basis. In addition, the harvest day (harv) and the initial amount of GLA on the day of emergence are estimated on a per-parcel basis in this study.

The parameters emerg, harv, SENa and GLA_init are directly deductible from the LAI curve, as described in Section 4.4.2. Deriving these parameters directly from the LAI curve reduces the processing cost of the calibration of the other parameters. The rest of the variable parameters which describe the phenology of the plant (SLAa, PRT_La, PRT_Le, SENb), as well as the efficiency of biomass production (LUEa), are derived through an optimization algorithm outlined in Section 4.4.3.

Initial parameter ranges for the calibrated parameters were defined using a ±2 standard deviation interval around the mean values reported by Pique et al. (2020b); see Table 2. However, for SENb the value range was visually refined using results from 20 randomly selected parcels to better capture phenological variation, because initially the parameter range was too narrow for the different barley senescence curves. Furthermore, PRT_Le is constrained to fall between 100 and the onset of senescence (SENa), ensuring biological realism, because no new leaves should form after the senescence begins.

All non-calibrated parameters are based on the documentation of the model which provided a-priori parameter values for wheat in France. Since wheat and barley, as well as the climate regime in Southern France and Northern Spain, are relatively similar, the a priori values were used as given, which was confirmed by the model developers. See Appendix A for the full list of parameters used by the model.

Table 2: Parameters used by SAFYE-CO2 that vary by parcel and are extracted from the LAI curve or through calibration.

Name	Notation	Unit	Value range	Method
Emergence day of the year	emerg	DOY	30 - 110	LAI curve extraction
Harvest day of the year	harv	DOY	150 - 250	LAI curve extraction
LAI on DOY of emergence	GLA_init	m2/m2	-	LAI curve extraction
Sum of temperature at DOY of LAI max	SENa	°C	-	LAI curve extraction
Rate of senescence	SENb	°C/day	50,000 - 350,000	Calibration
Amount of DAM that is not alocated to leaves at emergence	PRT_La		0.025- 0.625	Calibration
Sum of temperature at end of leaf allocation	PRT_Le	°C	100 - SENa	Calibration
Ratio of leaf area to leaf dry mass	SLAb	m2/g	0.006 - 0.014	Calibration

(Specific Leaf Area)				
Light Use Efficiency Parameter A	LUEa	gC/MJ	0.95 - 1.15	Calibration
Amount of DAM converted to grain	PRT_G		0.01 - 0.99	Calibration

4.4.2 Determining parcel-specific parameters from LAI curve

In the model, plant growth begins on the DOE (Day of Emergence), when GLA_init sets the initial amount of available leaves. The DOE is primarily identified by an increase in the LAI compared to a previously stable LAI. Based on the crop calendar, the DOE is assumed to take place between DOY 30-110. To detect the first increase in LAI, the slope (i.e., the first derivative of LAI with respect to time) and the slope change (i.e., the difference between successive slope values) are computed. The slope indicates whether LAI is increasing or decreasing, whereas the slope change indicates whether the rate of increase is increasing (>0) or slowing down (<0). In addition, a dynamic LAI threshold value is set to avoid overinterpreting subtle fluctuations in the LAI. This threshold is defined individually for each parcel to account for variability in baseline LAI levels. It is calculated as the mean LAI between DOY 30 and 60, plus an offset of 0.1, ensuring that only significant increases above the baseline are considered. Hence, the DOE is defined as the first date on which the slope and slope change are both positive, and the LAI exceeds the parcel-specific threshold. Figure 12 illustrates two examples for which the DOE was determined.

In addition, the LAI is extracted on the DOE and set as the initial GLA input to the model. This step ensures a realistic continuation of the curve, and avoids a delay in the modeled growth curve due to differing baseline LAI values.

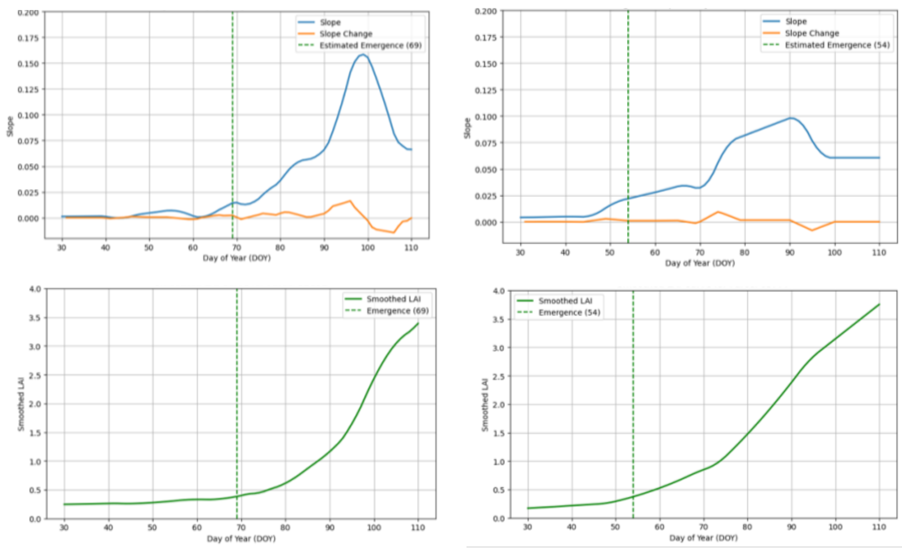


Figure 12: Calculation of the DOE for two parcels. The top graph displays the slope and change of slope, the bottom graph shows the LAI.

The Day of Harvest (DOH) marks the point at which all plant variables are reset to zero. In terms of the LAI, this is observed as the LAI declining and stabilizing at its baseline value. Accurate identification of the DOH is critical, as premature harvesting shortens the simulated senescence phase, which is essential for simulating yield development. The DOH is constrained to occur between DOY 150 and 250. Within this window, the DOH is defined as the first day when the LAI falls below 1, the slope change is negative—indicating a decreasing rate of LAI—and the slope on the subsequent day is smaller than on the current day. This last condition ensures the LAI decline is transitioning into a plateau, rather than continuing to decrease steeply. This method provides a robust balance between early and late DOH detection across different parcels. Figure 13 illustrates two examples for which the DOH was determined using this approach.

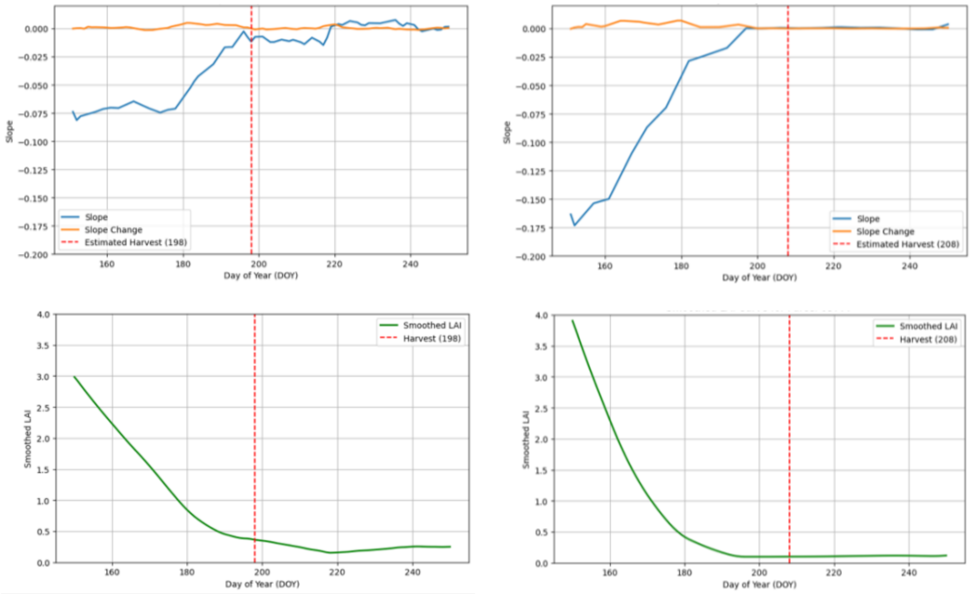


Figure 13: Calculation of the DOH for two parcels. The top graph displays the slope and change of slope, the bottom graph shows the LAI.

Lastly, the parameter SENa indicates when senescence begins, measured in SMT since the emergence. At this point, leaf growth transitions into leaf senescence, typically observed as the point when the LAI begins to decline after reaching the maximum LAI.

Two methods were used to estimate SENa. The first method identifies for each parcel the DOY corresponding to the maximum LAI and calculates the sum of efficient daily temperature between the DOE and the DOY of the maximum LAI. It is therefore max-LAI-based and henceforth termed SENa 1. The sum of efficient temperature is calculated as:

$$SMT = \sum_{i=DOE}^{DOY\ at\ LAI_{max}} T_{air,i} - T_{min,i} \qquad 11$$

The second method is based on Zhang et al. (2003) where the onset of senescence is based on the change in the curvature of the LAI curve, hence named slope-based SENa and referred to as SENa 2. This is particularly relevant if the LAI plateaus for a couple of days before starting to decline. Therefore, the second method calculates the slope of the LAI curve to identify the DOY where the slope is minimized. Between the DOE and the DOY of the minimum slope, the

SMT is calculated as in Equation 11. As displayed in Figure 14, the DOY of the minimum slope can occur multiple days after the maximum LAI is reached.

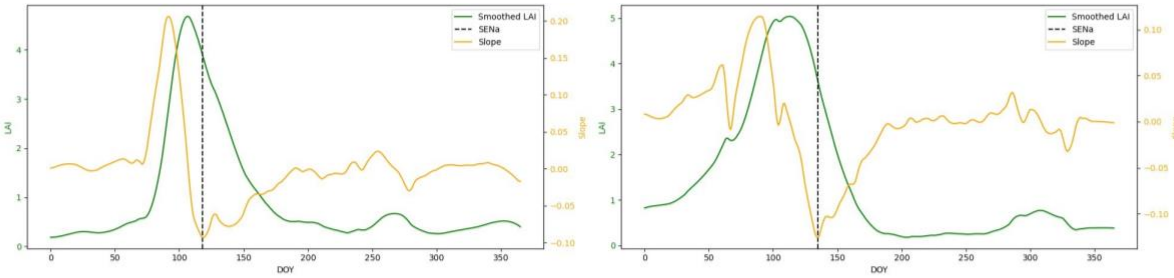


Figure 14: Calculating the DOY of SENa 2 for two parcels. The green line represents the LAI, the yellow line represents the slope of the LAI curve, and the dotted black line represents the DOY on which the slope reaches its minimum.

4.4.3 Selection of optimization algorithm for the calibration of parcel specific parameters

The aim of the calibration process is to identify a unique set of the vegetation parameters SLAb, PRT_La, PRT_Le, SENb and LUEa for each parcel that best represent the parcel's temporal dynamics of the LAI. To achieve this, an objective function is defined based on the Root Mean Square Error (RMSE) between the simulated and observed LAI values.

The optimization algorithm systematically explores the parameter space, defined in Table 2, by repeatedly running the model with different parameter sets to minimize the objective function, ultimately selecting the parameter set that yields the lowest RMSE.

One of the key challenges in such calibration procedures is the issue of equifinality, meaning the presence of multiple parameter sets that result in similar RMSE values (Mendez-Morales & Calvo-Valverde, 2019). This phenomenon often arises due to parameter co-dependence or when several parameters exert similar influences on model outputs. For instance, both SLAb and LUEa contribute to enhancing LAI growth during the vegetation period, while PRT_La constrains this growth. As a result, various combinations of these parameters may yield indistinguishable LAI trajectories. To mitigate the effects of equifinality and ensure more robust parameter estimation, careful selection of the calibration algorithm and a thorough sensitivity analysis of the parameters are essential.

To select the most suitable optimization algorithm, both local and global optimization strategies were considered. Local methods, such as gradient-based algorithms, can efficiently find optima when the objective function has a single minimum. However, they are highly sensitive to initial conditions and prone to becoming trapped in local minima, making them less suitable for problems characterized by complex parameter interactions (Mendez-Morales & Calvo-Valverde, 2019). In contrast, global optimization algorithms aim to explore the entire parameter space, minimizing the risk of settling in local optima (Mendez-Morales & Calvo-Valverde, 2019). Given the limited prior knowledge about optimal parameter values and their variability across parcels, global optimization was deemed more appropriate for this study. Three global approaches were evaluated: Grid Search, Latin Hypercube Sampling (LHS), and Bayesian Optimization (BO).

Grid Search is the most exhaustive method, evaluating all possible parameter combinations within defined bounds (Liashchynskyi & Liashchynskyi, 2019). Since this can quickly become computationally expensive, LHS is a more efficient alternative that divides each parameter's range into equally spaced intervals and samples once from each, ensuring a well-distributed coverage of the multidimensional space (Shields & Zhang, 2016). This approach dramatically reduces computational time while still maintaining broad exploration of the parameter space. Lastly, Bayesian Optimization is a probabilistic global optimization method that efficiently samples the parameter space through continuous learning, balancing exploration (sampling unknown areas) and exploitation (refining areas likely to contain the minimum) (Candelieri, 2021). The most time-efficient and accurate method was selected for implementation.

4.4.4 Sensitivity analysis

Due to sensitivity in the initialization of the calibration process (i.e., the first set of parameters used by the calibration algorithm) and the potential co-dependencies among parameters, a sensitivity analysis was conducted to better understand equifinality. As suggested by Candelieri (2021), the selected calibration algorithm was repeated 50 times on two randomly selected calibration parcels and then 15 times on 100 parcels to assess uncertainty in predictions across a larger set of parcels. Each run was initialized with a different set of parameter values. The spread of the resulting optimized parameters was evaluated alongside the output variables GLA and DGM. In a second step, to assess the sensitivity of the GLA and DGM to the parameters, a one-at-a-time sensitivity analysis was performed. Each parameter was varied in equal steps across the whole calibration range and the impact on GLA and DGM was recorded. This analysis served two purposes: (1) to identify whether different parameter sets lead to similar GLA outcomes yet different DGM outcomes, and (2) to evaluate the impact small variations in the parameters have on GLA and DGM.

4.5 Calibration of the Harvest Index and estimating yield

To determine the optimal HI value, the selected optimization algorithm was applied to the calibration parcels. Subsequently, the simulated grain yield per parcel was compared to the observed yield using RMSE. The HI was then fine-tuned using the scipy.optimize.minimize function with parameter bounds constrained between 0.0001 and 0.9999. The default optimization method, L-BFGS-B (Limited memory Broyden–Fletcher–Goldfarb–Shanno algorithm with bounds), was employed with an initial estimate of 0.5. The resulting HI was then applied to the validation parcels, to ensure consistency and avoid overfitting.

4.6 Restricting LAI assimilation for pre-harvest yield estimation

To investigate the model's ability to accurately calibrate the model parameters and predict the yield before the harvest takes place, the assimilation process was performed with LAI data until 1-2 months pre-harvest. Since the parcels are harvested at different times, they are divided into two groups: parcels harvested in June, and parcels harvested in July or later; see Figure 15.

For parcels harvested in June, the LAI data was restricted until 30.04.2020, whereas for July parcels LAI data was restricted until 31.05.2020.

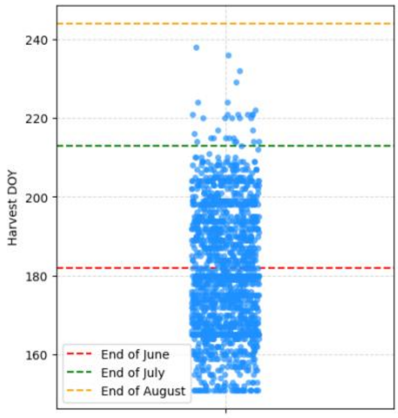


Figure 15: Distribution of harvest dates across rainfed, eye-estimated parcels (using SENa 1).

4.7 Sen4Stat Machine Learning Algorithm for yield estimation

To benchmark the performance of the process-based model against a statistical model, the Sen4Stat toolbox was employed for yield prediction. Sen4Stat requires only parcel boundaries and their corresponding yield values for model training. For each of the three parcel groups, the calibration dataset was provided to Sen4Stat, enabling it to extract 43 vegetation and meteorological features (Bontemps et al., 2024). Vegetation features are based on LAI values from the Sentinel-2 timeseries extracted at key dates, as well as from the SAFY model. Meteorological features are extracted from ERA5-Land precipitation, radiation, temperature, and soil water data (Bontemps et al., 2024). A machine learning regressor was then trained using these input features and the calibration dataset. In a second step, the trained model was applied to the validation parcels to generate yield estimates, allowing for a direct comparison with both observed data and outputs from the process-based models.

4.8 Using the ACEO LUT for yield estimation at pixel level

As presented in the literature review in Section 2.6.2, the SAFYE-CO2 model is integrated into the AgriCarbon-EO processing chain, as visualized in Figure 16. In particular the version 1.5.0, made available by CESBIO, is applied (Wijmer et al., 2019). To run ACEO, the user must provide (1) a shapefile with the parcels to be investigated, (2) a time period for the simulation, and (3) files containing soil and vegetation parameters.

ACEO first downloads weather data using the ERA5-Land hourly reanalysis dataset (1950–present) (see Section 3.2.4), and Level 2A satellite imagery from Sentinel-2. The Sentinel-2 imagery is then converted into pixel-wise LAI values using the PROSAIL model. PROSAIL is a radiative transfer model that couples the PROSPECT leaf optical properties model and the SAIL canopy reflectance model (Duan et al., 2014).

After processing the weather and LAI data, ACEO generates a Look-Up-Table for each ERA5-Land grid cell. The SAFYE-CO2 model is run multiple times per weather grid cell, using parameter sets sampled from the provided ranges of the eight parcel-specific parameters (see Table 2). Each run produces a simulated LAI curve and corresponding DAM, which are stored with the parameters in the LUT.

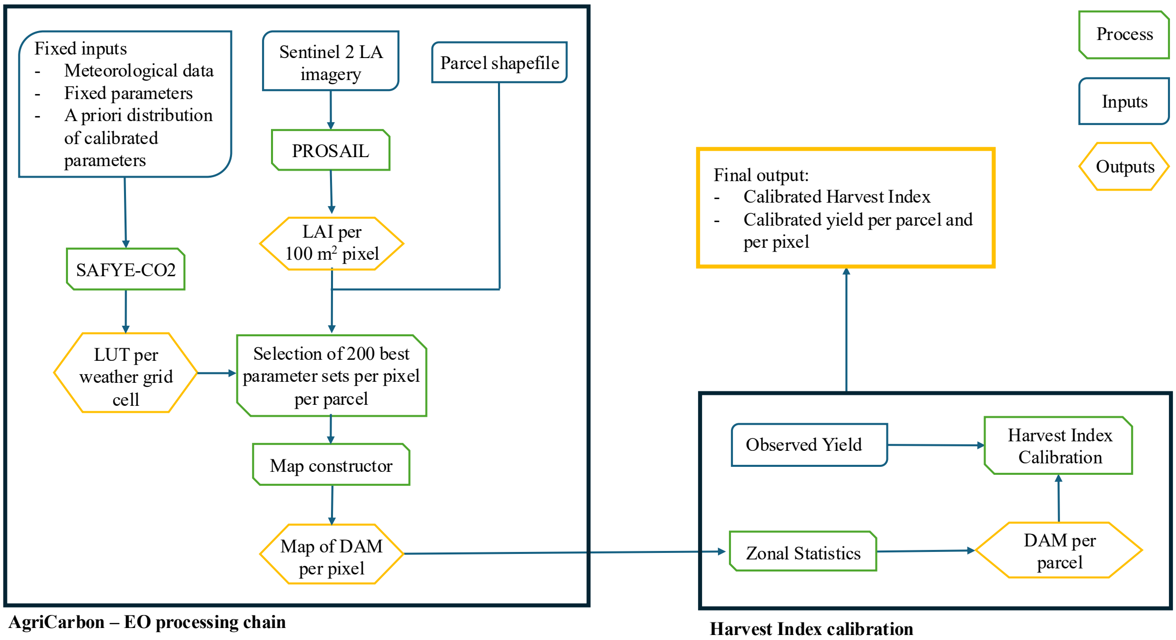


Figure 16: Workflow of the AgriCarbon-EO processing chain in combination with Harvest Index calibration for yield estimation at parcel level

Although generating the LUTs requires an upfront computational investment, typically involving thousands of simulations per weather grid cell, the approach is computationally efficient in the long term. Once the LUT is constructed, the evaluation of pixels is much faster, as it only involves comparing observed LAI to precomputed model outputs rather than rerunning the full model.

To estimate the DAM per pixel, the LAI retrieved via PROSAIL is compared to the modeled LAI stored in the LUT for the relevant grid cell. The best-matching LAI curves are identified, and their associated DAM outputs are averaged to provide mean and standard deviation for each pixel. The final output is a 10 m resolution GeoTIFF file, where each pixel contains a DAM estimate.

This contrasts with the direct modeling approach described in Section 4.4, where the model must be run repeatedly for each parcel (e.g., 150 simulations per parcel for calibration). In a single ERA5-Land grid cell, which covers ~9 km x 9 km, there may be over 100 parcels, which would result in more than 15,000 individual model runs. In contrast, the LUT approach requires only a single set of runs per weather cell, regardless of the number of parcels or pixels.

Furthermore, ACEO produces results at 100 m^2 pixel - not parcel - level. This enables a finer spatial resolution and allows the model to account for within-field heterogeneity, which would be very expensive with the direct modeling approach. A single parcel may contain hundreds of pixels. Hence, computing pixel-level outputs via repeated full-model runs would be computationally infeasible at this scale. Thus, the LUT-based method should offer both scalability and high-resolution insights into the spatial variability of crop performance.

Due to the high upfront processing, ACEO was tested on all barley parcels within one Sentinel-2 tile that covers a large part of the study area, including around 700 parcels with observed yield information (see Figure 2 for the location of the tile). The eight parcel-specific parameters, as previously defined in Table 2, were varied (except for PRT_G/HI) and the rest of the parameters were kept constant. For each weather cell, the number of parameter sets to be generated was set to 7000, of which the best 200 were selected for each pixel to calculate the mean and standard deviation of DAM.

To calculate the yield, the pixel-wise DAM was averaged per parcel and the HI had to be derived from the relationship between yield and DAM:

$$Yield_{parcel} = HI * DAM_{parcel} \qquad 12$$

The HI was calibrated on a subset of 300 parcels using the same method as described under Section 4.5.

4.9 Validation statistics

To quantitatively assess the accuracy of the modeled yield in comparison to the observed yield, six validation metrics are calculated, namely the RMSE, RRMSE, Mean Absolute Error (MAE), R^2, Nash-Sutcliffe Efficiency (NSE), the bias, and the Coefficient of Variation (CV).

4.9.1 Root Mean Square Error

Despite its critiques, the RMSE is a widely used metric to evaluate model performance (Chai & Draxler, 2014). It provides an indication of the average magnitude of error between the observed and modeled values. By squaring the errors, RMSE ensures that positive and negative errors do not cancel each other out, however, it gives a greater weight to larger errors. As a result, it is sensitive to outliers. The RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=0}^{n} (y_i - \hat{y}_i)^2} \qquad 13$$

Where y_i are the observed values, $\hat{y}_i$ the modeled values, and n the total number of observations.

4.9.2 Relative Root Mean Square Error

Since the RMSE is dependent on the scale of the evaluated variable, the RRMSE can be computed to represent the RMSE as a percentage of the mean observed values. This metric is particularly useful when comparing model performance across different crops or regions, where the mean yield values are different. By normalizing the error relative to the magnitude of the observed values, the RRMSE can be better compared across different datasets. The RRMSE is calculated as follows:

$$RRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=0}^{n} (y_i - \hat{y}_i)^2}}{\bar{y}} * 100 \qquad 14$$

Where y_i are the observed values, $\bar{y}$ the mean of the observed values, $\hat{y}_i$ the modeled values, and n the total number of observations.

4.9.3 Mean Absolute Error

An alternative to the RMSE is the MAE (Chai & Draxler, 2014). In contrast to the RMSE, it does not place greater weight on larger errors. Instead, it takes the absolute error value to avoid the cancellation of negative and positive errors. To ensure comparability to different studies, both RMSE and MAE are calculated for this study. The MAE is calculated as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad 15$$

Where y_i are the observed values, $\hat{y}_i$ the modeled values, and n the total number of observations.

4.9.4 Coefficient of Determination (R^2)

In addition to error metrics, assessing how well the model captures the variance in the observed data, is needed. The Coefficient of Determination (R^2) provides a measure of the variance explained by the model (Di Bucchianico, 2007). The closer the R^2 to 1 the better model performance, whereas values < 0.5 usually suggest a weak explanatory power of the model in relation to the observed variance. The R2 is calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - f(y_i))^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \qquad 16$$

Where y_i are the observed values, $\bar{y}$ the mean of the observed values, $f(y_i)$ the regression between the observed and modeled values, and n the total number of observations.

4.9.5 Nash-Sutcliffe Efficiency

Similar to R^2, the NSE is a metric to capture how well the model explains observed variations. However, while R^2 is based on the regression between the modeled and observed values, NSE is calculated relative to the identity line. The NSE is calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \qquad 17$$

Where y_i are the observed values, $\bar{y}$ the mean of the observed values, $\hat{y}_i$ the modeled values, and n the total number of observations.

4.9.6 Bias

The final metric used to evaluate the performance of the model is bias. The bias indicates whether there is a systematic over- or underestimation in the modeled values. A positive bias indicates consistent overestimation, while a negative bias indicates underestimation. Bias is calculated as:

$$Bias = \frac{\sum_{i=1}^{n} y_i - \hat{y}_i}{n} \qquad 18$$

Where y_i are the observed values, $\hat{y}_i$ the modeled values, and n the total number of observations.

4.9.7 Coefficient of variation

The CV, or relative standard deviation, is used to compare the parameter ranges resulting from multiple runs for one parcel with different initial conditions, to identify the issue of equifinality. The CV is calculated for each parameter per parcel as:

$$CV = \frac{\sigma}{\mu} \qquad 19$$

Where μ is the mean value for a parameter based on the different runs for a parcel, and σ its standard deviation.

5. Results

5.1 Calibration algorithm selection

To efficiently perform parameter calibration through LAI assimilation, three global optimization algorithms (Grid Search, Latin Hypercube Sampling, and Bayesian Optimization) were tested. Grid Search, though exhaustive, was computationally impractical for calibrating five continuous parameters, requiring multiple hours per parcel. Therefore, only LHS and BO were tested on four representative parcels, where BO consistently outperformed LHS in both speed and accuracy (see Table 3). Attempts to further refine solutions using the local optimization algorithm L-BFGS-B in combination with LHS or BO yielded no significant improvement in RMSE and were not pursued further (see Table 3).

Consequently, Bayesian Optimization was selected for calibrating the five continuous vegetation parameters. The number of parameter sets explored, that is the number of model runs, was set to 150, after which no improvements in the RMSE were observed (see Appendix B).

Table 3: Comparison of performance between Latin Hypercube Sampling (LHS) and Bayesian Optimization (BO) as well as in combination with a local gradient optimization method (L-BFGS-B) on four randomly selected parcels.

	RMSE between observed and modeled LAI			
Parcel ID	LHS	LHS + L-BFGS-B	BO	BO + L-BFGS-B
31762	0.354	0.29	0.22	0.21
32192	0.63	0.57	0.33	0.33
52701	0.21	0.21	0.22	0.22
63777	0.42	0.41	0.32	0.31

5.2 Sensitivity analysis

The impact of the one-at-a-time variation of the parameters on the maximum LAI, DOY of maximum LAI, and DGM is presented in Figure 17, Figure 19, and Figure 20.

Out of the five parameters to be calibrated, SENb does not affect the maximum LAI, as it primarily influences the rate of senescence (see Figure 18). In contrast, the light use efficiency parameter LUEa moderately impacts LAI max, contributing to an LAI variation of around 2. Higher LUEa values result in higher LAI max values. Both PRT_Le and SLAb strongly increase the LAI, spanning an LAI range of about 10. The parameter with the most pronounced effect is PRT_La, which causes an LAI range of almost 25, where higher values lead to a significantly lower LAI max values. Although these extreme LAI values are not realistic, they arise from unrealistic parameter combinations, which will be avoided through the assimilation of satellite-derived LAI. Nevertheless, this analysis clearly highlights that SLAb, PRT_La, and

PRT_Le exert the strongest influence on LAI max, where small variations in the parameters can lead to large variations in LAI max. Furthermore, it becomes apparent that SLAb and PRT_Le behave complementarily (both increase the LAI max), whereas PRT_La counteracts their effect. This finding supports the possibility of equifinality, meaning that multiple parameter combinations exist for similar LAI curves.

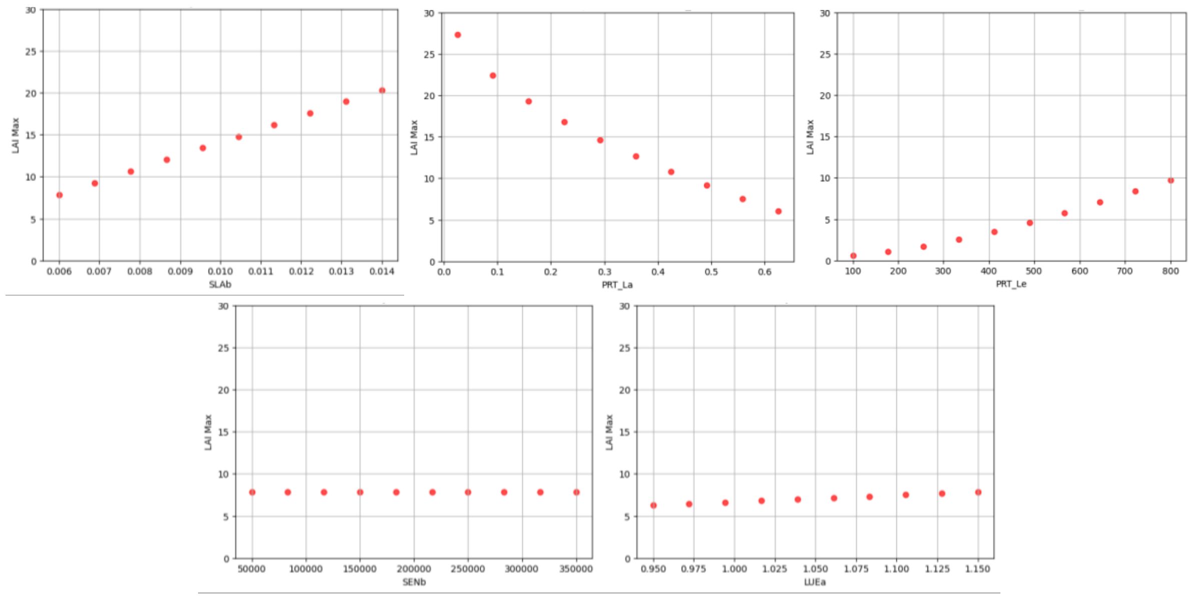


Figure 17: One-at-a-time sensitivity of maximum LAI to the 5 parameters to be calibrated. The constant parameters were set to the following value: SLAb = 0.006, PRT_La = 0.545, PRT_Le = 693, SENb = 160,315, LUEa = 0.95. These values were derived through Bayesian Optimization to best describe the parcel's LAI curve.

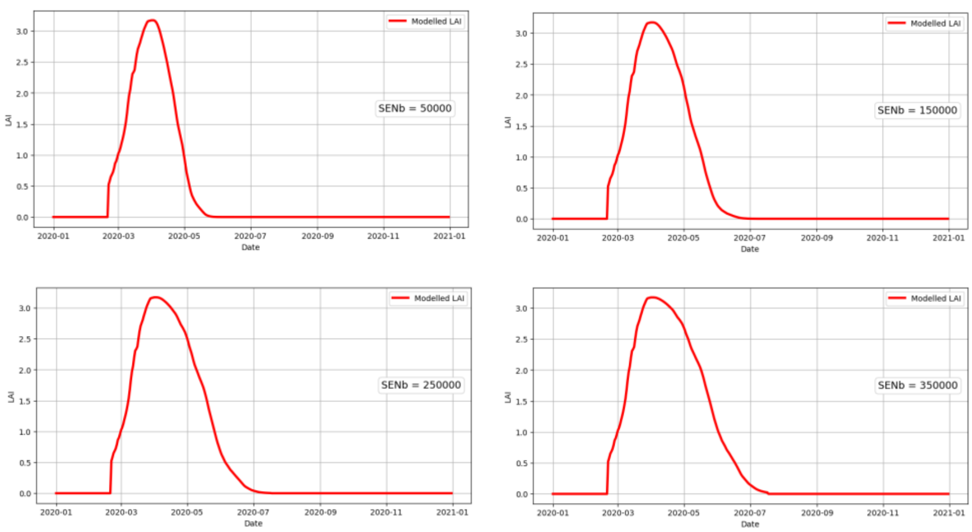


Figure 18: impact of SENb on the senescence displayed on an example aprcel by varying SENb.

As shown in Figure 19, the DOY of the LAI max is solely affected by PRT_Le. It indicates the day on which the leaf growth stops. Hence, it naturally affects the DOY of LAI max. The higher the PRT_Le, the later the DOY of LAI max.

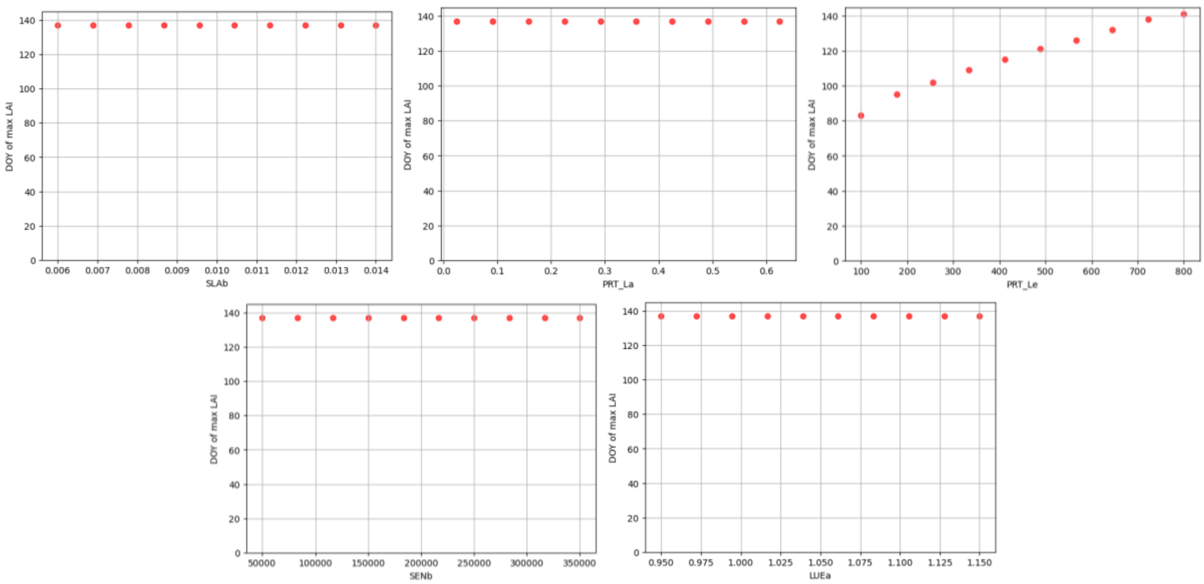


Figure 19: One-at-a-time sensitivity of Day od Year (DOY) of maximum LAI to the 5 parameters to be calibrated. The constant parameters were set to the following value: SLAb = 0.006, PRT_La = 0.545, PRT_Le = 693, SENb = 160,315, LUEa = 0.95. These values were derived through Bayesian Optimization to best describe the parcel's LAI curve.

Regarding yield, SLAb and PRT_La have only a marginal effect (see Figure 20). The most influential parameter is PRT_Le, which causes yield differences exceeding 5 tons per hectare. This can be attributed to the fact that if leaf growth stops early (low PRT_Le), the time for grain production is too short. The second most influential parameter is SENb, with a yield variation of nearly 3 tons per hectar. SENb determines the senescence curve; a higher SENb slows the rate of senescence, thereby extending the grain-filling period. Lastly, yield is also sensitive to LUEa, as it controls how efficiently the plant converts intercepted radiation into biomass.

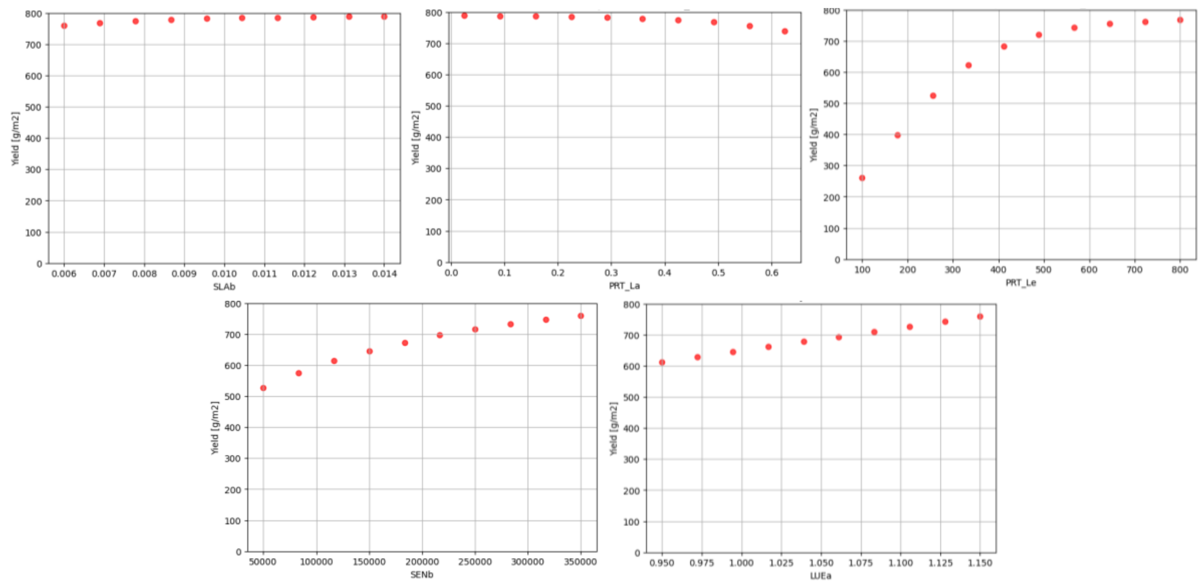


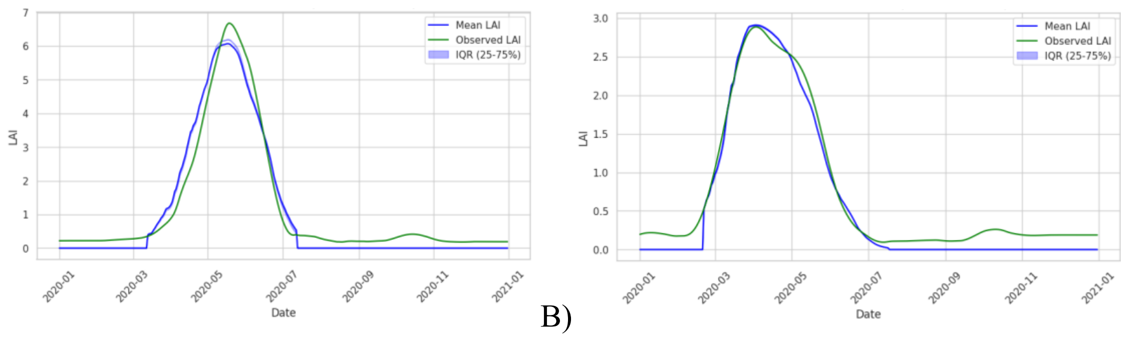
Figure 20: One-at-a-time sensitivity of yield [kg/ha] to the 5 parameters to be calibrated. The constant parameters were set to the following value: SLAb = 0.006, PRT_La = 0.545, PRT_Le = 693, SENb = 160,315, LUEa = 0.95. These values were derived through Bayesian Optimization to best describe the parcel's LAI curve.

Overall, the sensitivity analysis demonstrates that all five parameters play an important role in LAI and yield modeling. However, some parameters exhibit complementary effects—such as SLAb and PRT_Le, which both promote leaf growth—while others, like PRT_La, exhibit counteracting effects by restricting it. Furthermore, parameters such as LUEa have only subtle influences and might be poorly identifiable from the LAI curve. This poses a challenge for model calibration, as different parameter combinations can produce similar results. Such equifinality reduces the prediction capacity and reliability of the model. Therefore, the following section presents the results of the uncertainty analysis, which assesses the calibration's potential to identify unique parameter solutions.

5.3 Uncertainty analysis for equifinality identification

5.3.1 Investigation of individual parcels

To identify the uncertainty in parameter calibration and its effect on the LAI and yield, the Bayesian Optimization calibration process was run 50 times for two parcels that display unique LAI curves. As visible in Figure 21, each of the 50 calibration runs resulted in almost identical LAI curves that are very similar to the observed LAI. Consequently, the IQR overlaps with LAI curve and is barely visible. Only for parcel 18595, there was one extreme outlier that caused a shift in the interquartile range, likely the result of an impossible parameter combination as observed in the sensitivity analysis.



A) B)

Figure 21: Observed LAI curve, and the mean and interquartile range of the 50 modelled LAI curves for A) parcel 18595 and B) parcel 103610.

Figure 22 depicts the calibrated parameter values for each of the 50 runs. For both parcels, PRT_Le and SENb are clustering around one value. PRT_Le, which is restricted in its maximum value by the SENa value (i.e., the SMT on the DOY of LAI max), tends to move towards the maximum value (693 for parcel 18595 and 365 for parcel 103630). Since SENb is the only parameter affecting the LAI curve during the senescence phase, it can be correctly identified as it is not competing with other parameters. SLAb and PRT_La, on the other hand, both increase the LAI during the vegetative growth. Hence, they seem to balance each other out in both cases, as SLAb tends to be minimized and PRT_La maximized. They both show some clustering indicating that each calibration could produce a similar LAI curve and DGM. However, LUEa values spread across the whole calibration range. As shown in the sensitivity analysis, its effect on the LAI is much smaller than that of SLAb, PRT_La, and PRT_Le. Therefore, the variations of LUEa should have only a marginal impact on LAI curves. However, its effect on the yield was shown to range around 2 t/ha, which could negatively affect the accuracy of yield estimates and lead to high uncertainty.

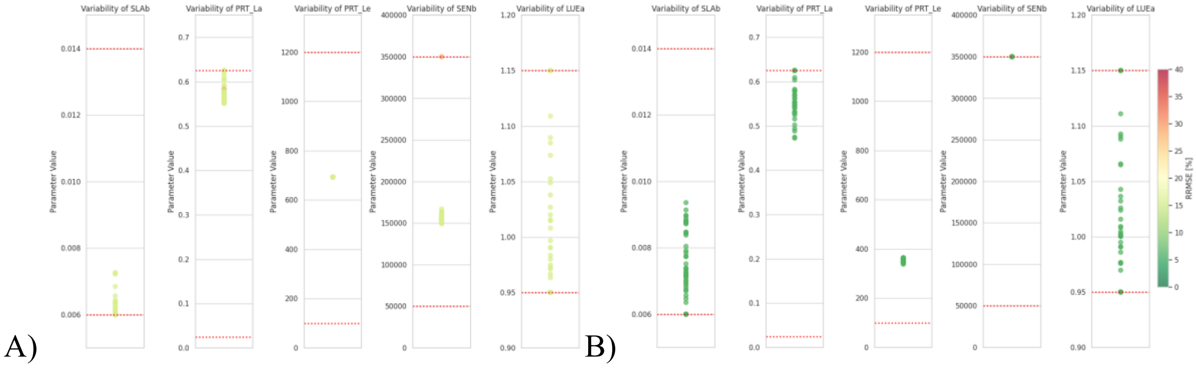


Figure 22: Calibrated parameter values for each of the 50 Bayesian Optimization calibrations for A) parcel 18595 with LAI max = 6.5 and B) 103610 with LAI max = 2.9. Colored based on the achieved RRMSE between modeled and observed LAI. A boxplot version can be found in the Appendix C.

Given this uncertainty, it was investigated whether setting SLAb or LUEa to a constant value, assigning PRT_Le = SENa, and only varying PRT_La, SENb, SLAb/LUEa would converge to one value. As shown in Figure 23, if LUEa was set, then the calibration found one minimum across all 50 runs. However, when SLAb was set, LUEa values still varied, suggesting that the model LAI is not sensitive enough to LUEa.

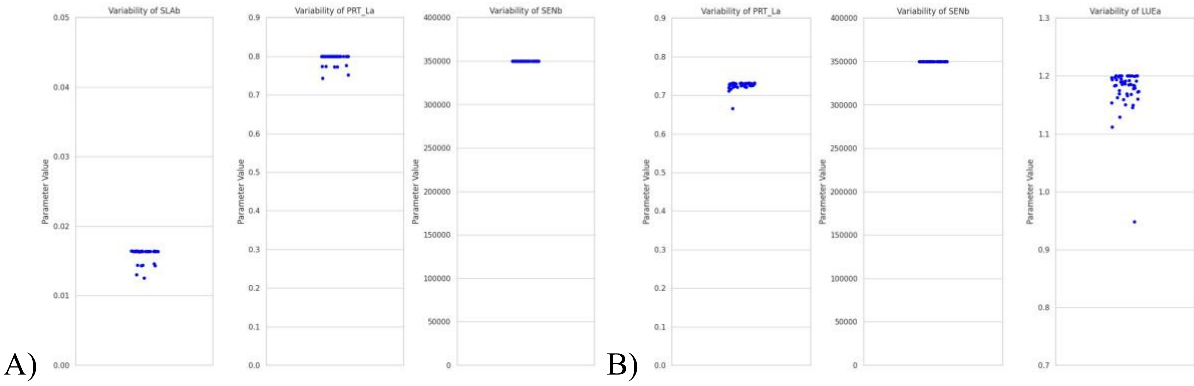


Figure 23: Parameter values after 50 Bayesian Optimization calibrations for parcel 103610 by A) varying only SLAb, PRT_La, and SENb, setting LUEa = 1, and PRT_Le = SENa; B) varying only PRT_La, SENb, and LUEa, SETTING SLAb = 0.01, and PRT_LE = SENa.

Despite the considerable variation in LUEa and moderate spread in SLAb and PRT_La, the resulting variability in simulated yield remains relatively small; approximately 1 t/ha, as visible in Figure 24. For parcels 18595 and 103610, the predicted yields cluster around 5 t/ha and 3.5 t/ha, respectively, closely aligning with the observed yields of 5 t/ha and 3.7 t/ha. This indicates that, although the calibration process is susceptible to equifinality, the model consistently converges toward realistic yield estimates across the 50 runs. The limited spread in yield despite parameter uncertainty suggests that the model exhibits a degree of robustness in simulating LAI and yield, even when multiple local optima are present in the parameter space.

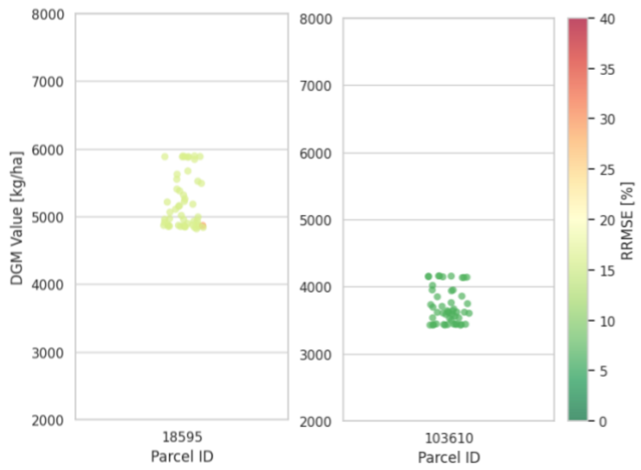


Figure 24: Final DGM value (i.e., yield) for each of the 50 individual runs per parcel. The RRMSE indicates the RRMSE between the modeled and observed LAI for each parcel run.

5.3.2 Generalization of the uncertainty

In order to generalize the uncertainty across a wider range of parcels, 50 of the calibration parcels were selected to be run 15 times, each run with different initial conditions. For each of the parcels, the coefficient of variation and the total range spanned by the 15 runs are assessed for the parameters and the yield. The results are displayed in Figure 25 and Figure 26.

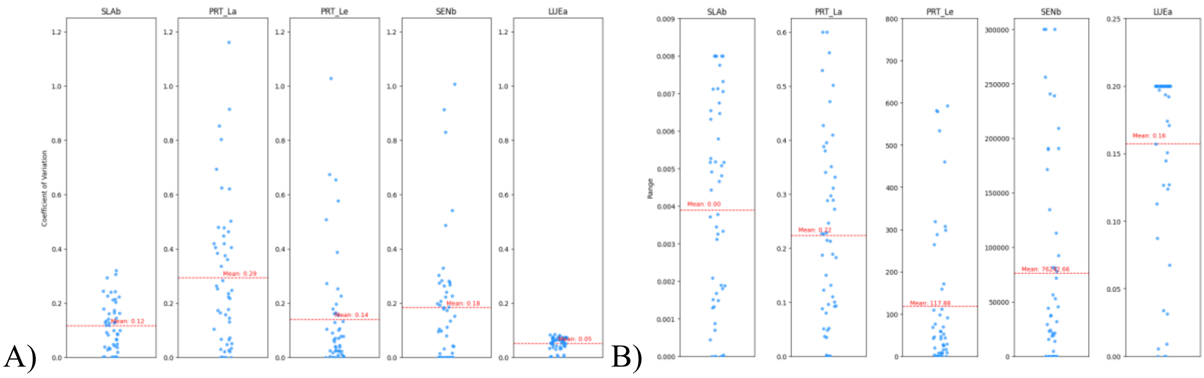


Figure 25: A) The coefficient of variation and B) the range, i.e. difference between maximum and minimum parameter value, achieved through 15 calibration runs by using different initial parameter sets. Each point represents one parcel result, in total 50 parcels were investigated.

Considering the CV, the parameter LUEa scores the lowest, with an average of 0.05, meaning that per parcel the calibrations vary approximately 5% around the mean. SLAb, PRT_Le and SENb score lower values as well with a variation between 12% and 18% around the mean. Less optimal conditions with a variation of almost ⅓ around the mean are observed for PRT_La. Furthermore, there are outliers of CV >1 for PRT_La, PRT_Le, and SENb, which suggests that the optimization was unable to consistently converge to one value for some parcels.

However, looking at the range, i.e., the difference between the minimum and maximum value achieved for a parameter across all runs for one parcel, offers additional insights. If the range value across the 15 runs for one parcel is close to 0, then for each run a similar parameter value is achieved. If the range value is close to the maximum range, then different runs for the same parcel lead to different parameter outcomes. The maximum range, i.e., the difference between the restricted bounds, for each parameter is SLAb = 0.008, PRT_La = 0.6, PRT_Le ≈ 600, SENb = 300,000, LUEa = 0.2.

Figure 25B indicates that despite the low CV, the parameters SLAb, PRT_La and LUEa span a substantial portion—or even the entire possible range—of their predefined calibration bounds for at least half of the parcels. As shown in the sensitivity analysis, variations in SLAb and PRT_La have a strong influence on LAI, however, their effects appear to counterbalance one another, with SLAb promoting and PRT_La limiting leaf growth. Their impact on yield is rather small. In contrast, LUEa spans the whole range likely because the LAI is relatively insensitive to it and instead is driven by SLAb, PRT_La, and PRT_Le. Nonetheless, LUEa's effect on the yield is relevant, as the sensitivity analysis showed its potential to cause a variation of almost 2 t/ha.

In comparison, the parameters PRT_Le and SENb appear more sensitive and well-constrained during calibration. Most parcels have a range of less than 100 and 50,000 for PRT_Le and SENb, respectively and many cluster around zero, which means that for these parcels the calibration consistently converges on the exact same value. Looking at the sensitivity analysis, uncertainty in the calibration of these two parameters does not translate into high uncertainty in the yield as long as the absolute values are at the upper end, i.e., above 400 for PRT_Le and 200,000 for SENb. However, as displayed in Figure 26, most of the parcels have a low PRT_Le and SENb, meaning that a change of 100 for PRT_Le and 50,000 for SENb could cause a difference in yield between 1-2 t/ha. Nonetheless, most parcels have very little variation in these two parameters, indicating that they can be effectively identified in the calibration process.

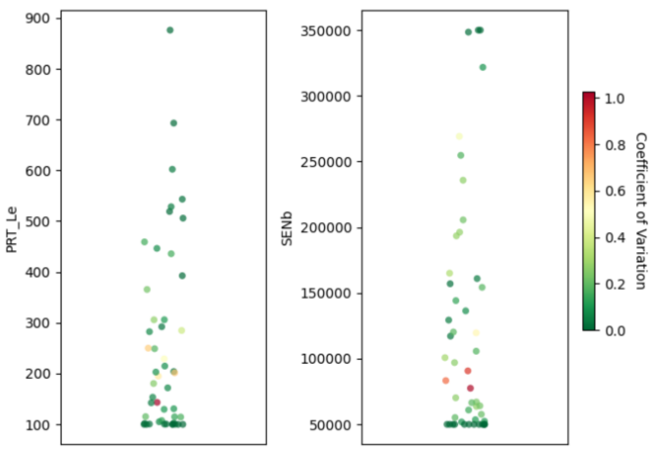


Figure 26: Distribution of PRT_Le and SENb values for 50 calibration parcels. The values are the mean of the 15 runs, and colored based on their coefficient of variation.

The effect of the calibration process on the yield is displayed in Figure 27. It shows the range for 50 parcels, based on their 15 calibration runs. The calibration results in an average yield range of almost 1 t/ha per parcel. While half of the parcels have a variation of less than 1 t/ha, there is also a significant number of parcels with a higher variation. Considering that the average yield is 4 t/ha, an uncertainty of 1 t/ha caused by the calibration is not desirable as it is about ¼ of the total yield.

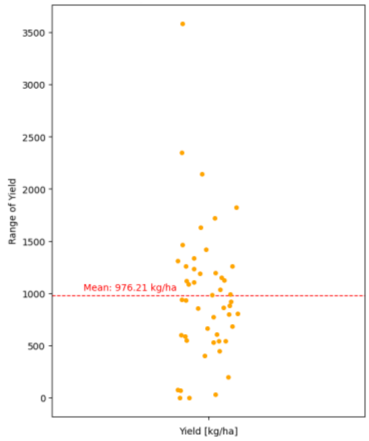


Figure 27: Range of yield for each of the 50 parcels, based on 15 Bayesian Optimization Calibrations per parcel.

Overall, the uncertainty analysis continues to support the finding that the parameters PRT_La and SLAb have a counteracting effect, which causes a high variability in their calibration, yet their effects on the yield and LAI are low. The LUEa, although being deemed an important parameter in previous literature (Duchemin et al., 2008; Pique et al., 2020a; Pique et al., 2020b), has a low sensitivity to the model and LAI yet a moderate effect on the yield. Nonetheless, in the following steps the calibration and yield estimation are performed using the five parameters with the ranges as defined above while acknowledging an uncertainty in yield of around 1 t/ha due to the calibration process.

5.4 Harvest Index calibration and validation

While the previously discussed parameters can be calibrated based on the LAI curve, the harvest index, referred to as PRT_G in the model, determines the amount of biomass allocated to the grain and therefore needs to be calibrated using observed yield values. The HI was calibrated separately for each of the three groups and both SENa methods (max-LAI-based and slope-based), using the calibration parcel sets. The resulting HI per group, along with the corresponding R^2 and RMSE between modeled and observed yield are displayed in Table 4. For all groups, the R^2 is low, indicating poor relationship between the observed and modeled yield. Nonetheless, no large deviations are observed between the calibration and validation parcel sets, suggesting that the calibrated HI values generalize well to parcels not included in the calibration and can be applied.

Only for irrigated parcels the validation parcels display a better performance in RMSE than (but worse R^2) than the corresponding calibration parcels. This indicates that although the absolute prediction errors are reduced for the validation parcels—suggesting closer agreement between predicted and observed yields—the model still struggles to capture the variability and trends in the data, as reflected in the lower R^2 and negative NSE values. This discrepancy can be attributed to the narrower yield range observed in the validation parcels (approximately 3,000–7,000 kg/ha), compared to the broader range in the calibration parcels (approximately 2,000–8,000 kg/ha).

Additionally, a clear difference between the SENa methods is notable, with SENa 1 having higher HI than SENa 2.

Table 4: Calibrated Harvest Index for 3 parcel groups and corresponding RMSE and R^2 for calibration and validation parcels.

			Calibration Parcels		Validation Parcels	
	Parcel set	Harvest Index	RMSE [kg/ha]	R^2	RMSE [kg/ha]	R^2
Max-LAI-based SENa	Eye-estimated	0.55	1151.7	0.198	1263.9	0.173
	Non-eye-estimated	0.57	1156.4	0.242	1293.2	0.176
	Irrigated	0.53	1671.3	0.145	1521.7	0.092
Slope-based SENa	Eye-estimated	0.46	1207.36	0.157	1546.4	0.126
	Non-eye-estimated	0.45	1251.04	0.19	1422.23	0.094
	Irrigated	0.46	1514.9	0.2	1318.3	0.158

5.5 Model evaluation at parcel level

5.5.1 Distribution of the observed yield

Before looking at the statistical evaluation, it is useful to examine the distribution of the observed yield in relation to the maximum LAI, as displayed in Figure 28. It becomes evident that there is some bias in the observed yield, likely due to rounding of yield estimates, resulting in horizontal lines in the observations. Moreover, except for the irrigated parcels, no clear linear relationship between maximum LAI and yield is visible, which would be expected, since higher LAI usually correlates with higher yield. Instead, a saturation effect can be observed, as the yield balances out at an LAI of 3 and a yield of 6 to 7 t/ha for (non-) eye-estimated parcels.

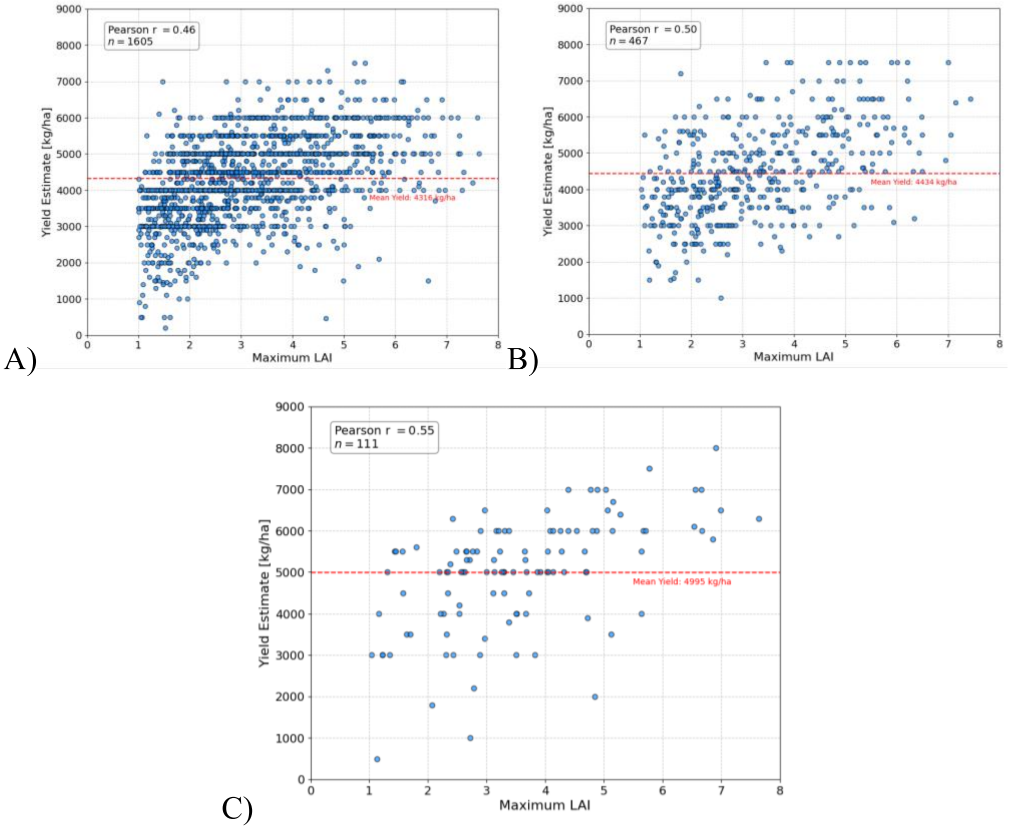


Figure 28: Distribution of observed yield in relation to observed LAI for A) eye-estimated parcels; B) non-eye-estimated parcels; C) irrigated parcels.

5.5.2 Eye-estimated yield

When examining the distribution of yield in comparison to the maximum observed LAI, a slight linear relationship between LAI and yield is observable (see Figure 29). However, at higher LAI the yield values become more dispersed and appears to saturate, in particular in Figure 29A. This can be a result of less accurate LAI assimilation, as indicated by the lighter colors, i.e., higher RMSE. In particular, for the SENa 2 method, the RMSE is higher, indicating a poorer assimilation performance.

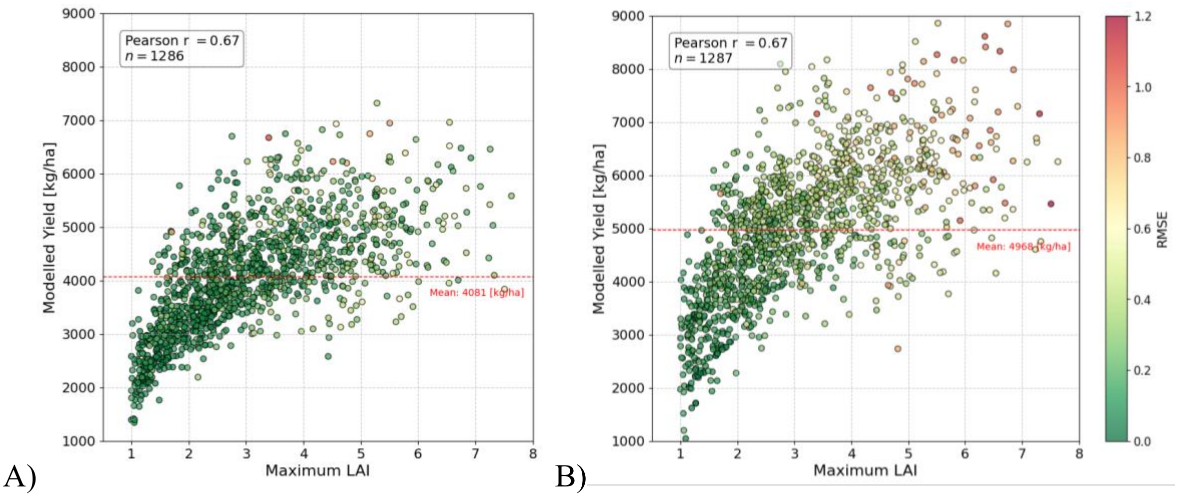


Figure 29: Distribution of modeled yield in relation to observed LAI, for eye-estimated parcels using A) SENa 1 and B) SENa 2. The RMSE is calculated as the difference between modeled and observed LAI.

As shown in Figure 30, comparing the modeled to the observed yield for SENa 1 (SENa 2) results in RMSE = 1192.25 (1510.20) kg/ha, RRMSE = 27.72 (35.11) %, MAE = 959.35 (1200.86) kg/ha, R^2 = 0.166 (0.139), NSE = -0.12 (-0.80), and a bias of -220.78 (666.87) kg/ha.

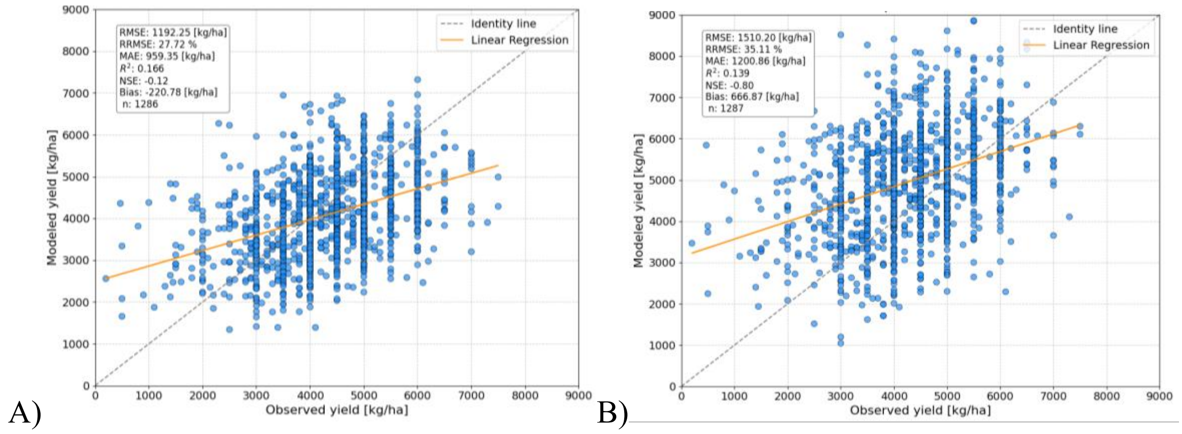


Figure 30: Comparison of observed to modeled yield for eye-estimated parcels using A) SENa 1 and B) SENa 2.

Considering the RMSE and MAE, the result is as expected based on the uncertainty analysis. However, both R^2 and NSE have poor values, indicating that the pattern and trend of the observed yield are not well captured by the model. This is also reflected in the bias, which shows a tendency of the model to underestimate the yield for SENa 1 (overestimate for SENa 2), yet it is smaller than the total error, as there are both over- and underestimations. For all metrics, the SENa 1 method performs better than the SENa 2 method.

5.5.3 Non-eye-estimated yield

Figure 31 shows the distribution of the modeled yield in relation to the maximum observed LAI for non-eye-estimated parcels. Again, a slight linear relationship is observable with a spread and slight saturation effect at higher LAI and yield values. Furthermore, the LAI assimilation performs better using SENa 1 and for lower LAI values, as indicated by the lower RMSE values.

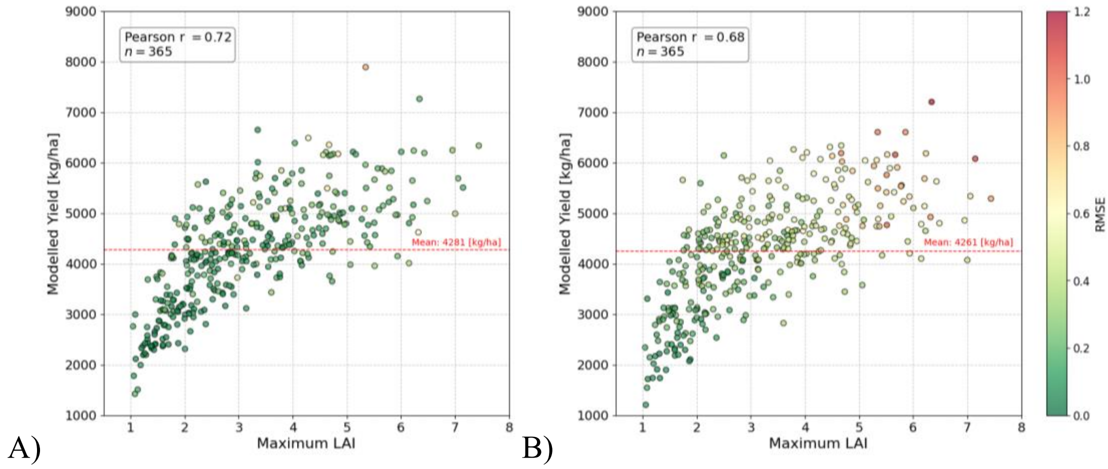


Figure 31: Distribution of modeled yield in relation to observed LAI, for non-eye-estimated parcels using A) SENa 1 and B) SENa 2. The RMSE is calculated as the difference between modeled and observed LAI.

As shown in Figure 32, the result is similar to that of eye-estimated parcels with SENa 1 performing better than SENa 2, with an error slightly above 1 t/ha, a negative bias, and a low R^2, indicating poor representation of the trends. The R^2 is slightly higher than the R^2 for eye-estimated parcels, which could be attributed to the less strongly visible bias in the observed yield. More precisely, the RMSE = 1238.95 (1342.42) kg/ha, RRMSE = 27.84 (30.17) %, MAE = 976.32 (1095.64) kg/ha, R^2 = 0.213 (0.134), NSE = 0.06 (-0.10), and bias = -169.03 (-189.36) kg/ha.

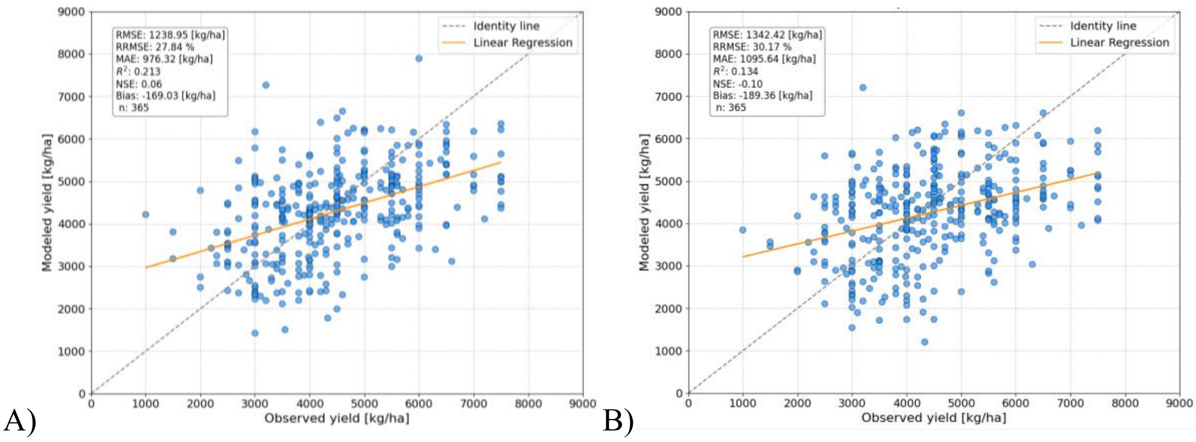


Figure 32: Comparison of observed to modeled yield for non-eye-estimated parcels using A) SENa 1 and B) SENa 2.

5.5.4 Irrigated yield

For irrigated parcels, the yield distribution in relation to LAI shows that the LAI assimilation is worse than for the other two groups, with most parcels having an RMSE of 0.4 or higher, see Figure 33. Yet, a similar linear trend with more variation at higher ends is visible.

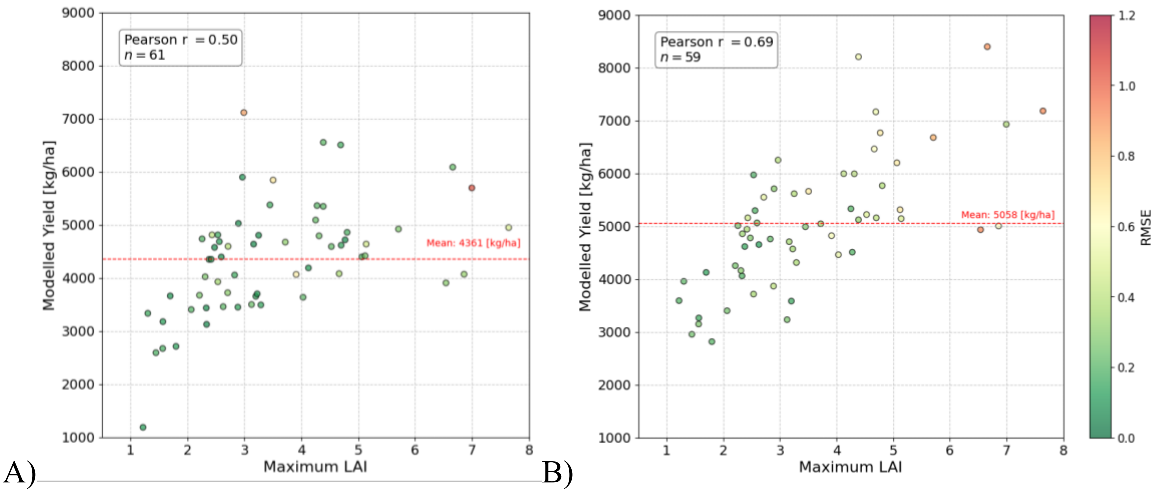


Figure 33: Distribution of modeled yield in relation to observed LAI, for irrigated parcels using A) SENa 1 and B) SENa 2. The RMSE is calculated as the difference between modeled and observed LAI.

Comparing modeled and estimated yield, the results are similar to the previous two groups, leading to an RMSE = 1512.71 (1318.33) kg/ha, RRMSE = 29.52 (25.73) %, MAE = 1290.47 (1031.51) kg/ha, R^2 = 0.092 (0.158), NSE = -0.65 (-0.21), and bias = -763.30 (-66.21) kg/ha (see Figure 34). Notably, for irrigated parcels the SENa 2 method yields better results.

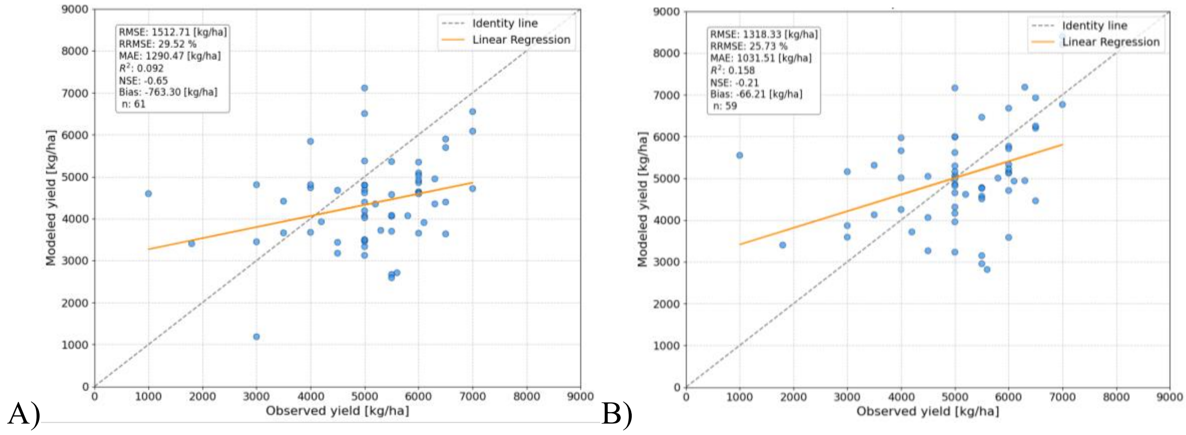


Figure 34: Comparison of observed to modeled yield for irrigated parcels using A) SENa 1 and B) SENa 2.

5.5.5 Comparison of max-LAI-based and slope-based SENa

Across all three parcel groups, the HI is consistently lower when using SENa 2, indicating that the modeled biomass and consequently yield is higher under this approach. However, SENa 2 yields poorer performance in LAI assimilation, as evidenced by higher RMSE values, and performs slightly worse for (non-) eye-estimated parcels in terms of yield estimation, yet better for irrigated parcels. To understand these differences, the influence of SENa on LAI and yield, as well as the resulting calibrated parameters, is analyzed below.

As shown in Figure 35, SENa has minimal impact on maximum LAI. Consequently, parameters which govern LAI growth, like SLAb, LUEa and PRT_La, remain largely unaffected, see Figure 36. Similarly, the day of year (DOY) corresponding to maximum LAI shows little variation between approaches, as shown in Figure 35. Although SENa is used as the upper boundary for calibrating the PRT_Le parameter, its values remain mostly unchanged. Hence, the growth phase of the LAI curve is equally well simulated across methods.

The main contributing factor to the difference in yield is SENa itself. With slope-based detection, SENa occurs later in the season, which delays the onset of senescence. As seen in the sensitivity analysis (Figure 35), this can contribute to a difference of multiple tons of DGM. A later senescence onset allows for prolonged biomass accumulation, thereby increasing yield estimates.

The slightly worse overall performance of SENa 2 may stem from changes in SENb. Under the slope-based approach, SENb exhibits reduced variability, with values clustering at the extremes. This limited variation may reflect a less accurate representation of the senescence curve and slightly affect the yield by producing less variability between the individual parcels.

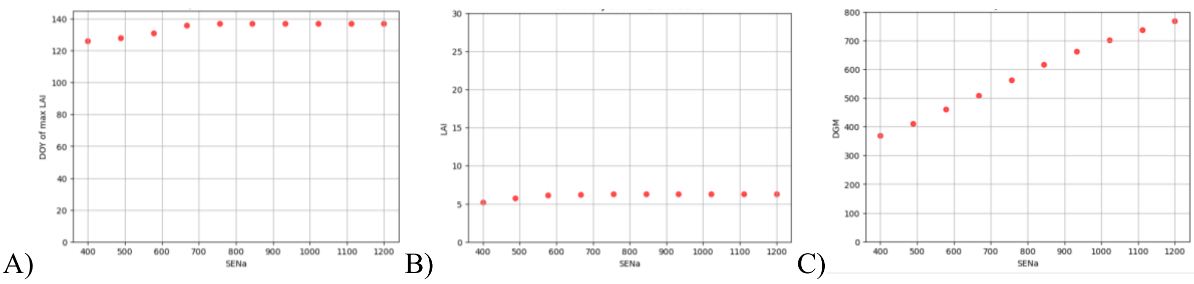


Figure 35: Sensitivity of A) DOY of LAI max, B) LAI max, and C) DGM to variation in SENa. The calibration parameters were set to the following value: SLAb = 0.006, PRT_La = 0.545, PRT_Le = 693, SENb = 160,315, LUEa = 0.95.

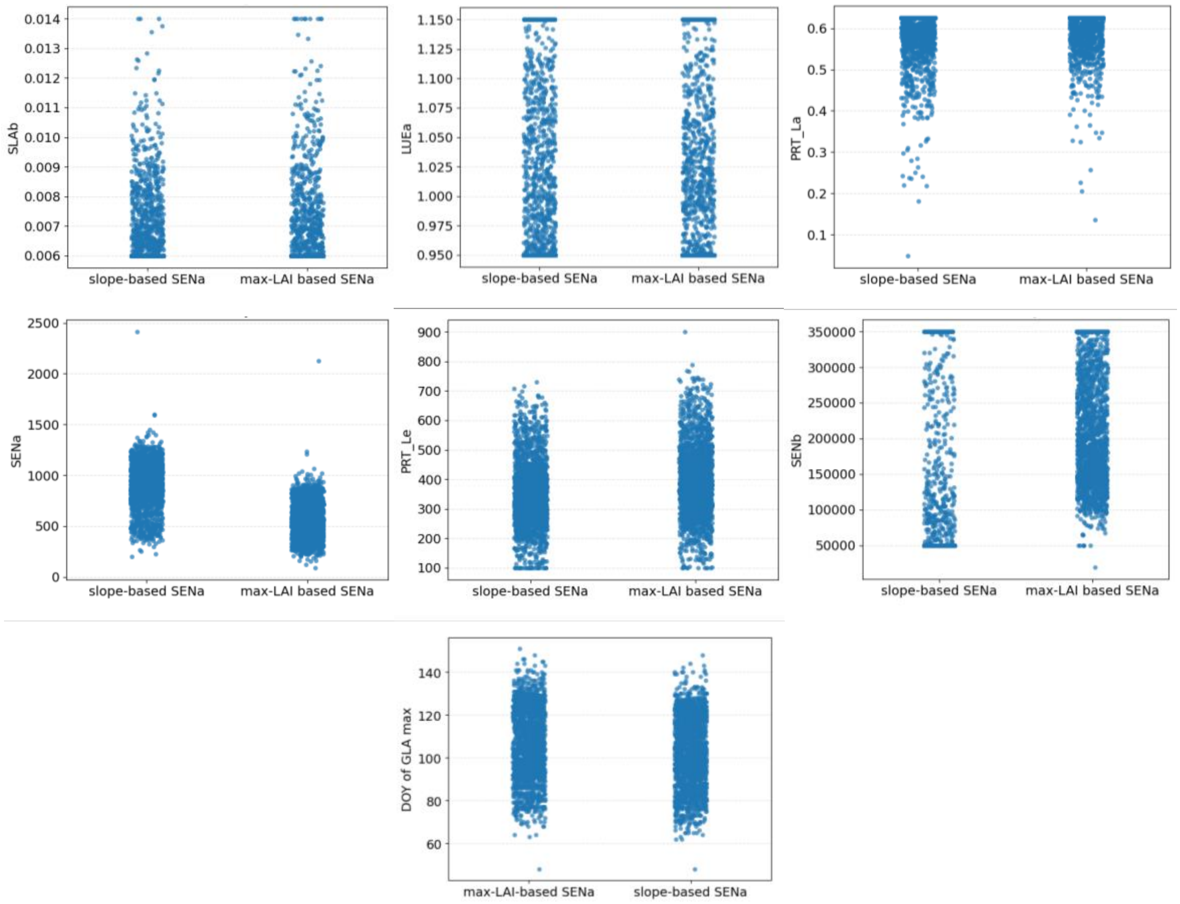


Figure 36: Distribution of calibrated parameters when using SENa 1(max-LAI-based) vs SENa 2 (slope-based).

Overall, the results at parcel level are not satisfactory, and contrary to the initial hypothesis the model performs similarly across all three parcel groups. Furthermore, the differences between the SENa 1 and SENa 2 methods are mainly adjusted for through the HI calibration. However, the yield observations were primarily collected to support provincial, regional, and national statistical reporting. Therefore, the next section presents the results after aggregating the modeled and observed yield per province, which aligns more closely with the intended scale of the observational data.

5.6 Model evaluation at provincial level

Given the finding that the SENa 1 method performs similarly, yet better than slope-based modeling, this section will only present provincial level results for the former. Results for the SENa 2 method can be found in Appendix D. Furthermore, due to the low number of irrigated parcels, the analysis focused on rainfed parcels.

As the parcels were aggregated based on their location, understanding the spatial distribution and density of parcels across provinces is crucial for interpreting the results. Figure 37 shows the number of parcels for each province in CyL. The most eye-estimated parcels are located in Valladolid (380), followed by Burgos (286) and Palencia (250). León has the fewest parcels counting only 20, and the rest of the provinces have between 100 and 150 parcels. For non-eye-estimated parcels, most parcels are located in Burgos (139), followed by Soria (99), and Segovia (98). The remaining provinces count between 20 and 30 parcels.

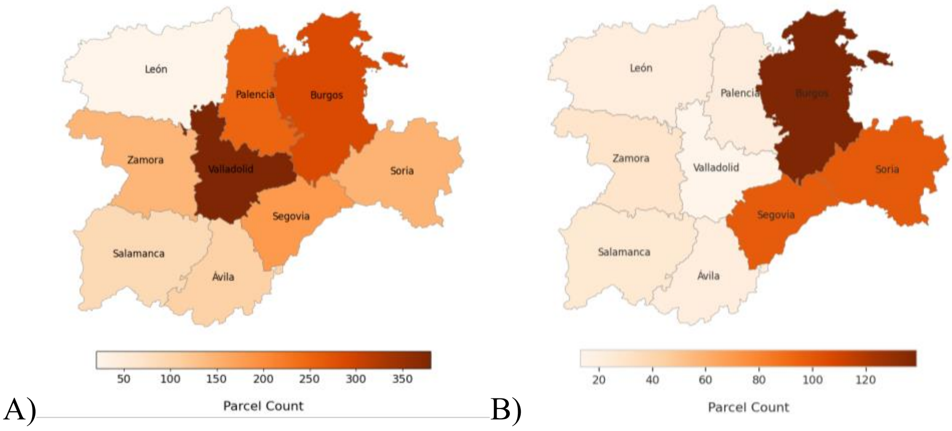


Figure 37: Number of Parcels per province for A) eye-estimated and B) non-eye-estimated parcels.

5.6.1 Eye-estimated parcels

As shown in the bar graph of Figure 38, the provinces with the highest number of parcels have the highest average yield, whereas the provinces with the lowest yields have fewer parcels. This pattern can likely be attributed to the fact that farmers have a higher incentive to grow barley in areas with higher barley yield than in areas with lower yield. Furthermore, the bar graph indicates that for provinces with higher yields, the model underestimates the yield, but for provinces with lower yields, i.e., Soria and León, the underestimation is not as strong, and even overestimates it for Soria.

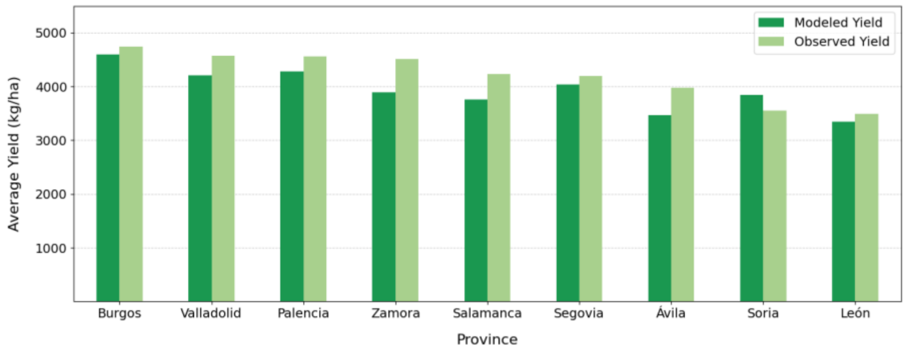


Figure 38: Comparison of observed and modeled yield at provincial level for eye-estimated parcels using SENa 1.

At provincial level, the R^2 and NSE have increased to 0.65 and 0.26, respectively. This indicates that the model has a good capability of capturing the trend in yield between different provinces. In addition, the RMSE, MAE, and RRMSE have been significantly reduced to 366.26 kg/ha, 328.71 kg/ha and 8.72%, respectively, meaning that the average error is only 8.72% of the average observed yield value, suggesting that the modeled values are relatively close to the observed values. However, given the bias of -264.72 kg/ha, the model tends to underestimate the yield, which is visible in both the scatterplot (Figure 39) and the bar graph (Figure 38).

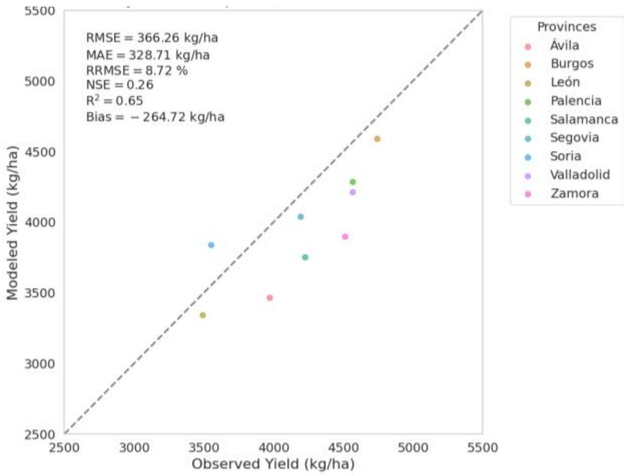


Figure 39: Comparison of observed and modeled yield at provincial level for eye-estimated parcels using SENa 1.

One explanation of the underestimation bias could be the calibration of the HI. Since the relationship between the modeled and observed yield is poor at parcel level, determining a HI is difficult, as indicated by the low R^2 value for the calibration parcels, lending to low explanatory power. Additionally, the parameter calibration could have influenced the underestimation bias, in particular the LUEa. As the LAI is not very sensitive to the LAI assimilation, it could be possible that most parcels were assigned a low LUEa value, which causes slightly lower yields. This is confirmed by Figure 40, which displays the distribution of LUEa values for all parcels. Most parcels are assigned a value between 0.95 and 1, causing slightly lower yields than parcels that were assigned a higher LUEa value. Many parcels are still assigned a LUEa value > 1, which is in line with the per-parcel analysis which clearly shows that many parcels have a higher modeled yield than observed. Yet, the overall bias to underestimate the yield could potentially be caused by the LAI's insensitivity to LUEa.

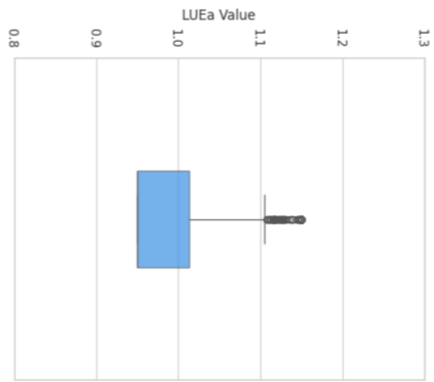


Figure 40: Distribution of calibrated LUEa values for eye-estimated parcels using SENa 1, showing that the majority of LUEa values falls below 1.

5.6.2 Non-eye-estimated parcels

As shown in Figure 41, the yield distribution per province is similar to that of eye-estimated parcels with the highest yields in Burgos and Valladolid, and the lowest yields in Soria and León. For this parcel group, for most provinces, the yield was overestimated, in particular for provinces with lower yields, except for Ávila, where the model largely underestimated the yield. This impacts the results in Figure 42. While the bias was decreased to -53.29 kg/ha, and lower RMSE of 386.79 kg/ha, which is about 9.21 % of the average yield, the R^2 has not improved significantly ($R^2 = 0.36$). This is caused by one outlier parcel in the Ávila province, since the rest of the provinces show a good agreement, see Appendix E.

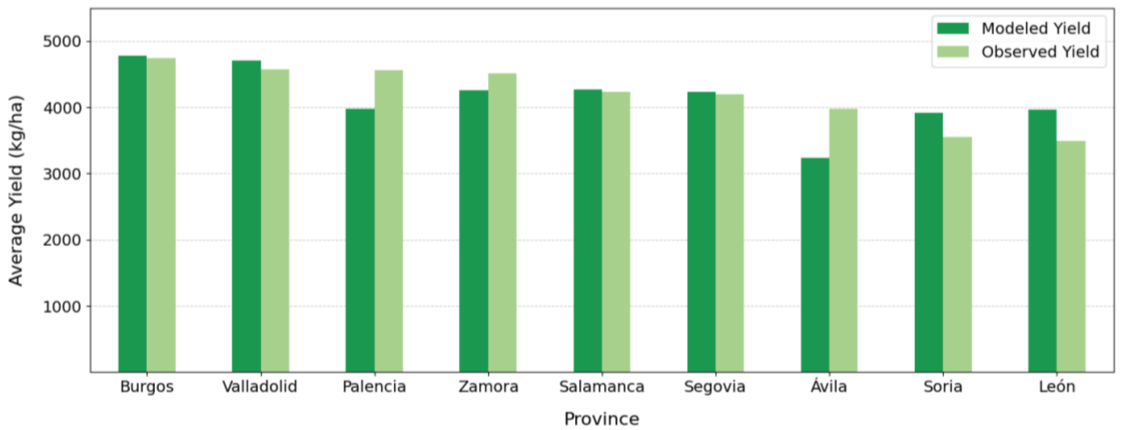


Figure 41: Comparison of observed and modeled yield at provincial level for non-eye-estimated parcels using SENa 1.

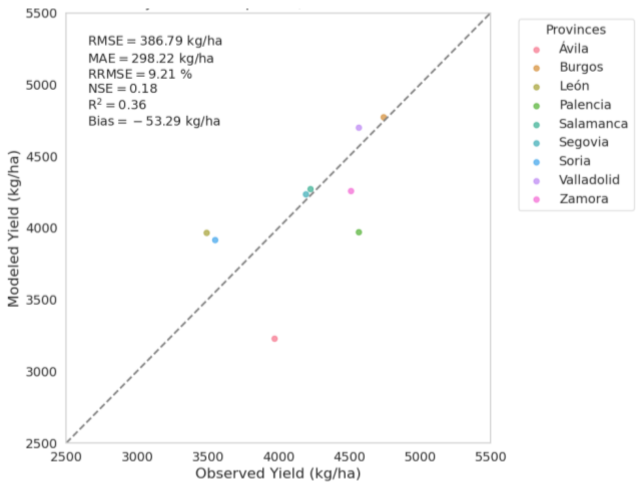


Figure 42: Comparison of observed and modeled yield at provincial level for non-eye-estimated parcels using SENa 1.

5.7 Yield estimation with restricted LAI data

In addition to estimating the yield at the end of the season, i.e., once the full LAI dataset is available, the ability to forecast the yield before the end of the season is a needed skill. In the following, the model's capability of estimating yield by using a restricted LAI dataset, i.e., until 1 to 2 months pre-harvest, is presented. The eye-estimated parcels are divided into two groups, i.e., June and July harvest.

On a parcel level, the predictability only slightly decreased, as displayed in Figure 43, with RMSE = 1312.62 kg/ha, RRMSE = 30.52% (compared to 27.72% when using the full LAI dataset), MAE = 1068.66 kg/ha, R^2 = 0.172, NSE = -0.36, and bias = -186.7 kg/ha. Looking at individual DGM curves for full (top) and restricted (bottom) LAI assimilation (Figure 44), the similar performance is confirmed, since in particular the curve shape is well simulated, i.e., whether it is flat or pointy.

Yet, while the values do not deviate much from the results when using the full LAI dataset, the model's performance remains poor, especially as indicated by the R^2 and NSE. The model fails to capture the variability and patterns of the observed yield for individual parcels.

Furthermore, a clear difference between June and July parcels is noticeable, with June parcels having a stronger negative bias of -505.7 kg/ha, i.e., a systematic underestimation, whereas July parcels show a systematic overestimation with a positive bias of 275.7 kg/ha.

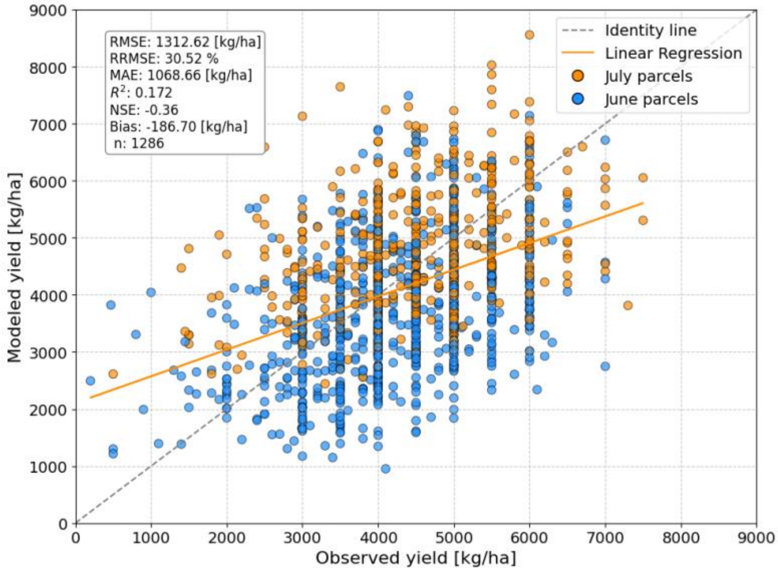


Figure 43: Comparison of observed to modeled yield for eye-estimated parcels using SENa 1. The LAI assimilation was restricted until 31.04.2020 for parcels harvested in June (blue points) and until 31.05.2020 for parcels harvested in July (orange points).

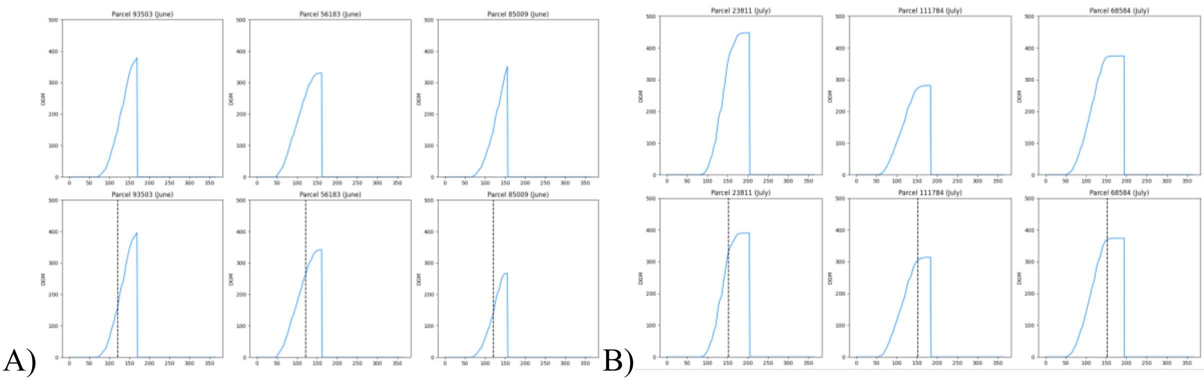


Figure 44: Three example DGM curves for A) June and B) July harvest, using full (top) and restricted (bottom) LAI assimilation.

However, at provincial level, the prediction error significantly decreases, see Figure 45A, and the model performs better than when using the full LAI dataset: RMSE = 332.21 kg/ha, RRMSE = 7.91%, MAE = 268.91 kg/ha, R^2 = 0.68, NSE = 0.39, and bias = -226.02 kg/ha. The RMSE again significantly decreased compared to the parcel level results and constitutes less than 10% of the average observed yield, indicating small discrepancies between observations and estimates. This is also reflected in the R^2 which indicates that around 68% of the variability in the data pattern is captured by the model. However, again a systematic underestimation bias is observed, which is also visible in Figure 46. The bar graph shows that the underestimation is taking place for almost all provinces, regardless of the average yield and number of parcels per province.

However, when looking at the June and July groups individually, the individual performance is not as good as the combined, see Figure 45B and 45C. For the June harvest group, the yield is underestimated, whereas for the July harvest group the yield is overestimated. Hence, the overall provincial result is likely due to over- and underestimation errors balancing each other out.

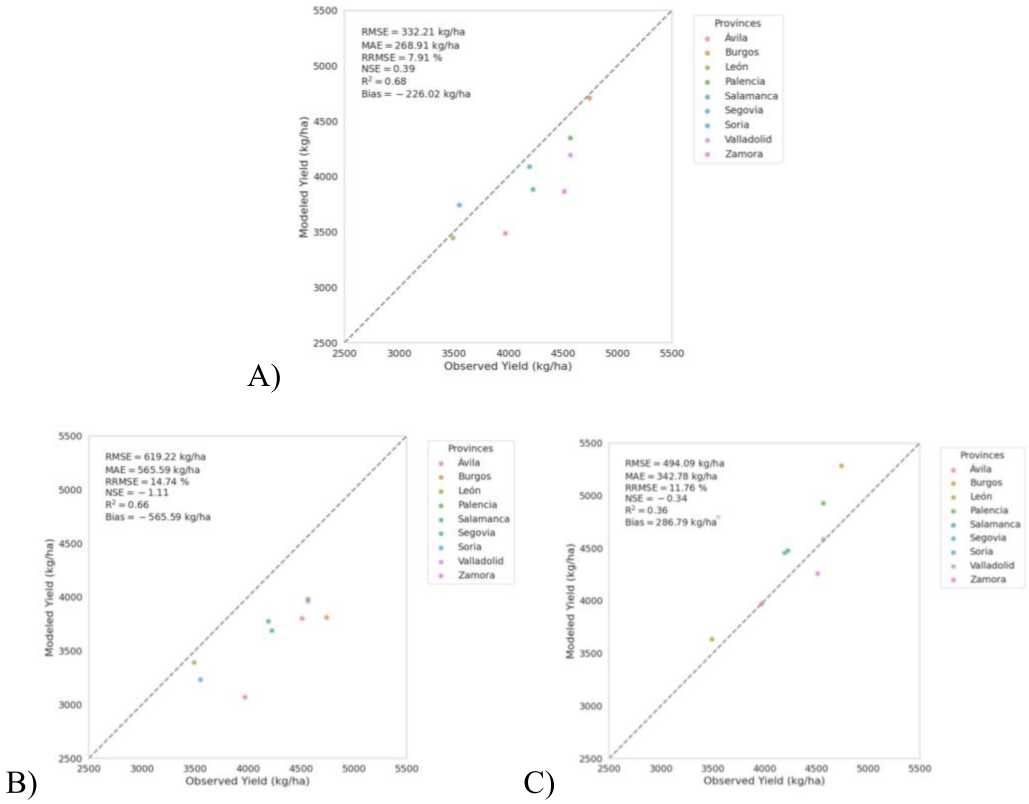


Figure 45: Statistical comparison of observed and modeled yield at provincial level for eye-estimated parcels using SENa 1 and restricted LAI assimilation. Graph A) displays both June and July harvest together, B) June and C) July parcels.

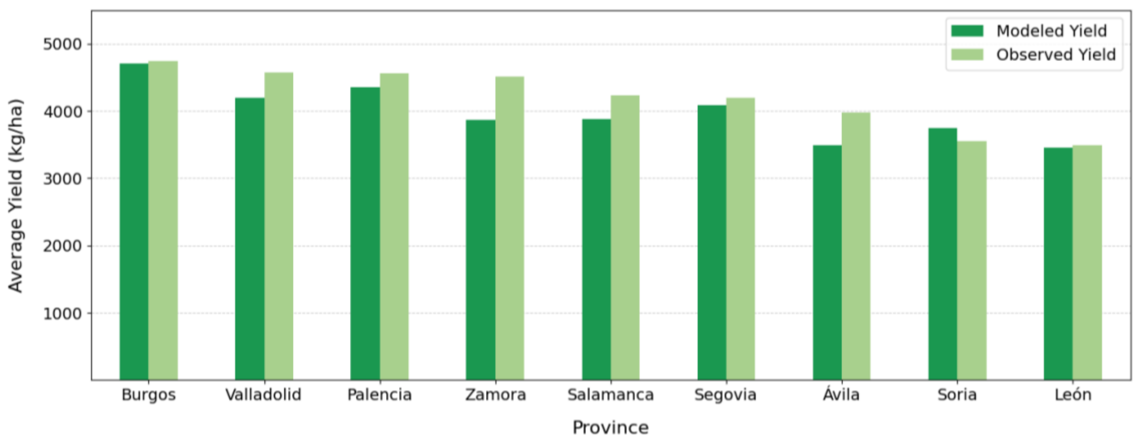


Figure 46: Comparison of observed and modeled yield at provincial level for eye-estimated parcels using SENa 1 and restricted LAI assimilation.

To better understand the observed differences between parcels harvested in June and those harvested in July under restricted LAI assimilation, it is important to consider the timing of maximum LAI and whether the assimilation period extends up to that point. This, along with the distribution of calibrated parameters, is examined below.

For parcels harvested in June, approximately half had not yet reached their observed maximum LAI by the time LAI assimilation was terminated. In contrast, most July-harvested parcels had already reached their peak, as shown in Figure 47A. Consequently, the restricted assimilation period led to an earlier modeled date of maximum LAI, particularly in the June group (Figure

47B). On average, the modeled LAI peak occurred 16 days earlier than observed for June-harvested parcels (IQR: 5 to 25 days), compared to 13 days earlier for July-harvested parcels (IQR: 3 to 21 days). An earlier simulation of the LAI peak leads to an earlier cessation of growth and lower modeled yield values, explaining to some parts the stronger underestimation bias observed for June-harvested parcels.

To further investigate this discrepancy, the distributions of the calibrated parameters are analyzed. As shown in Figure 48, most parameters exhibit similar distributions across both harvest groups. However, a notable difference appears for the parameter PRT_Le, which governs the end of leaf development. This difference could be partially attributed to the earlier harvest timing of June parcels, which naturally leads to an earlier end of the growth period, as well as the earlier LAI peak simulation. This also impacts the total simulated yield, leading to an underestimation in June parcels and overestimation in July parcels.

Additionally, the distribution of the SENb parameter reveals a greater number of parcels having low SENb values in the June group. Since lower SENb values accelerate senescence, this could also contribute to lower simulated yields, further explaining the tendency for underestimation of June-harvested parcels and overestimation of July-harvested parcels.

Overall, although there are differences between the June and July groups, mainly due to the capturing of the peak LAI, the restricted LAI assimilation still performs comparably to the full LAI assimilation approach.

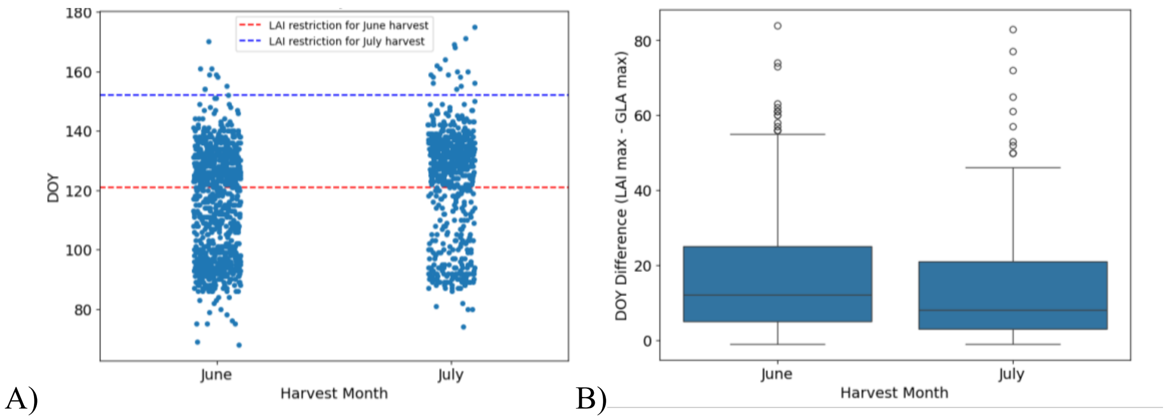


Figure 47: Distribution of A) DOY of LAI max, with lines indicating until when LAI assimilation took place and B) for the difference between observed and modeled LAI max, with positive values indicating that the modeled DOY of max LAI took place before the actual observed DOY of max LAI.

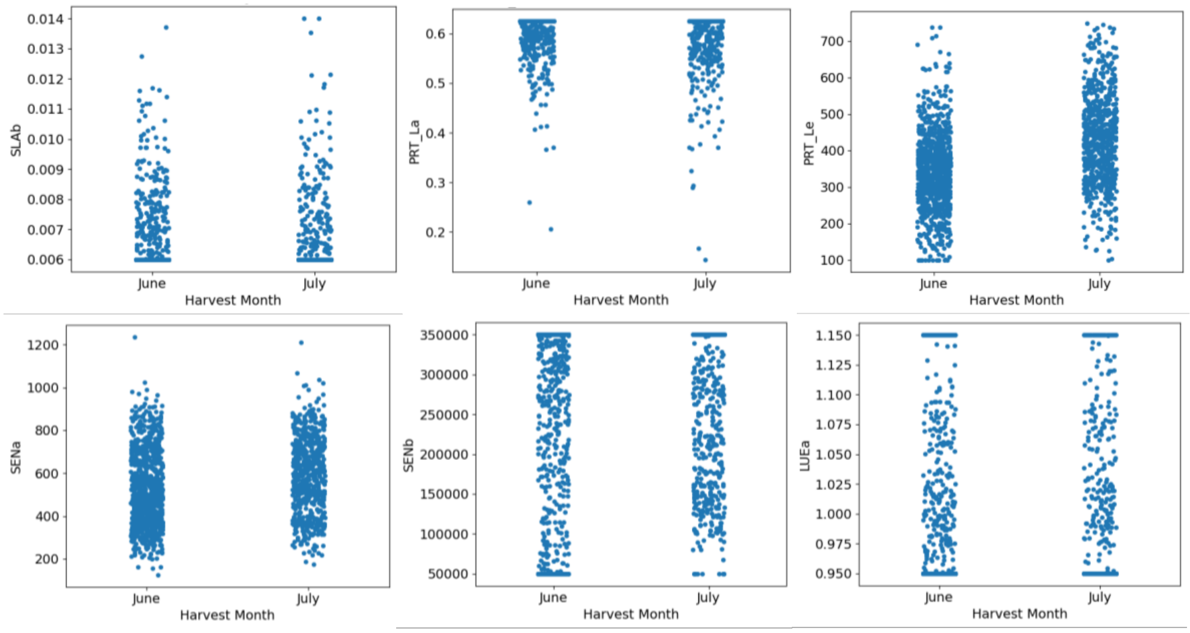


Figure 48: Comparison between calibrated parameters for June and July harvested parcels.

5.8 Yield estimates using the ACEO processing chain

Using the ACEO processing chain, instead of calibrating each parcel or pixel, a large set of simulated LAI curves and corresponding DAM is generated. By comparing the Sentinel-2-derived LAI curves to the modeled LAI curves, the best LAI curves are selected using BASALT.

The resulting DAM is at pixel level. By aggregating the values to parcel level and calibrating the HI between DAM and yield, resulted in a comparatively high HI of 0.77. Applying the HI to each DAM pixel a map with pixe-level yield values as in Figure 49 is created.

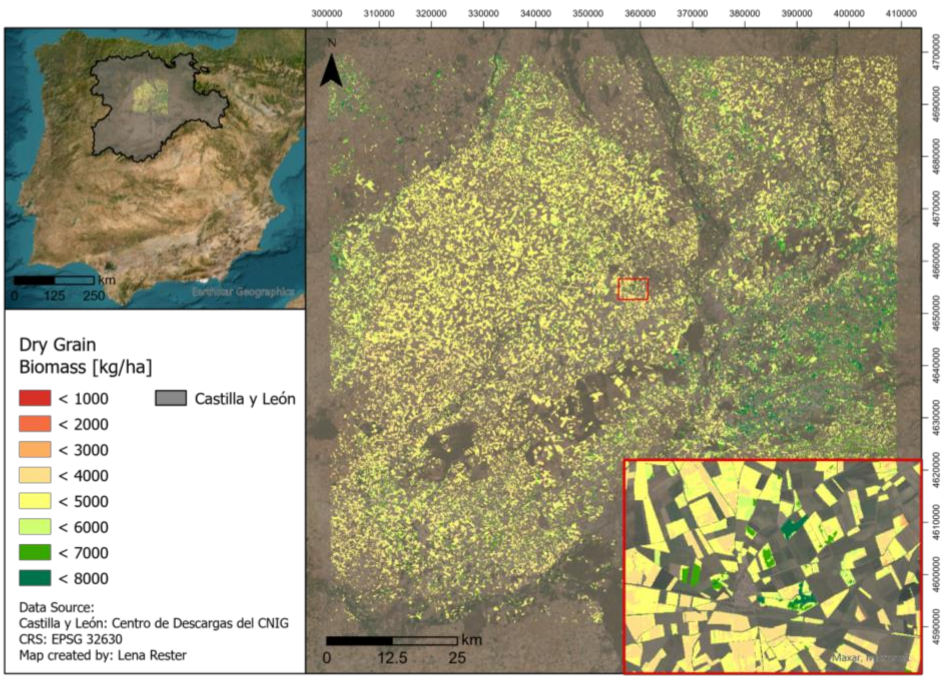
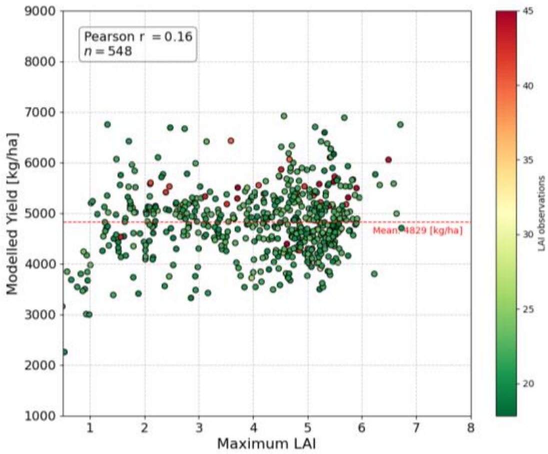


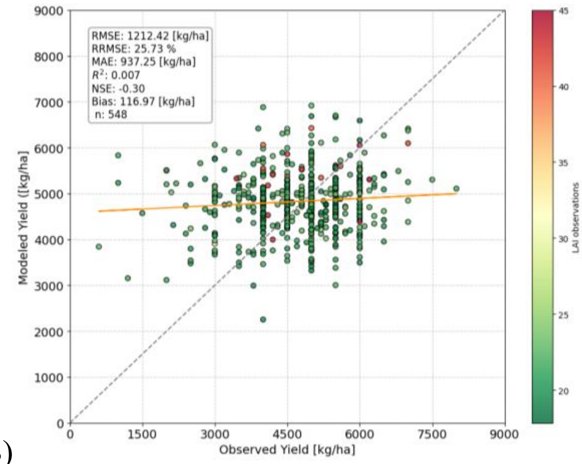
Figure 49: Pixel-wise yield estimates for barley parcels by applying the ACEO to one Sentinel-2 tile (T30TUM).

Comparing the modeled yield to PROSAIL LAI values and observed yield, as displayed in Figure 50A, shows that the ACEO has a very poor performance and is largely insensitive to the LAI values. This results in a horizontal spread of the yield values across the LAI values and observed yield. As quantified in the R^2, the ACEO has almost no explanatory power for the observed yield distribution, although producing a similar RMSE of 1212.42 kg/ha.

Next to the DAM results, the ACEO also provides information on the LAI data used. As shown in Figure 50B, most parcels had around 20 LAI values that could be used to extract LAI curves and corresponding yield values from the LUT.



A)



B)

Figure 50: Comparison of modeled barley yield using ACEO to A) PROSAIL LAI and B) observed yield. The color bar indicates the number of LAI acquisitions available for each parcel and to be used to identify most suitable modeled LAI.

5.9 Yield estimates using the Sen4Stat toolbox

In contrast to the ACEO and individual parcel calibration, using the Sen4Stat toolbox produces better results at parcel level. Figure 51 displays the modeled yield values for the validation datasets for the three groups. The best performance is achieved by eye-estimated parcels with RMSE = 996.3 kg/ha, RRMSE = 22.5%, MAE = 791.67 kg/ha, R^2 = 0.372, NSE = 0.36, and bias = 10.35 kg/ha. This is followed by the non-eye-estimated parcels with RMSE = 1016.76 kg/ha, RRMSE = 23.1%, MAE = 827.69 kg/ha, R^2 = 0.348, NSE = 0.33, and bias = -152.85 kg/ha. The poorest performance is again found for irrigated parcels with RMSE = 1182.38 kg/ha, RRMSE = 23.1%, MAE = 971.84 kg/ha, R^2 = 0.285, NSE = -0.01, and bias = -635.22 kg/ha.

Compared to the best-performing SAFYE-CO2 runs (and using only the validation parcels, for which the results are included in the Appendix F), this approach achieves an improvement of approximately 6–7 percentage points in RRMSE and increases the R^2 by 0.17 to 0.2. This is reflected in a tighter clustering of points around the regression line. However, the bias remains comparable—or even slightly more negative—indicating a persistent underestimation for non-eye-estimated and irrigated parcels.

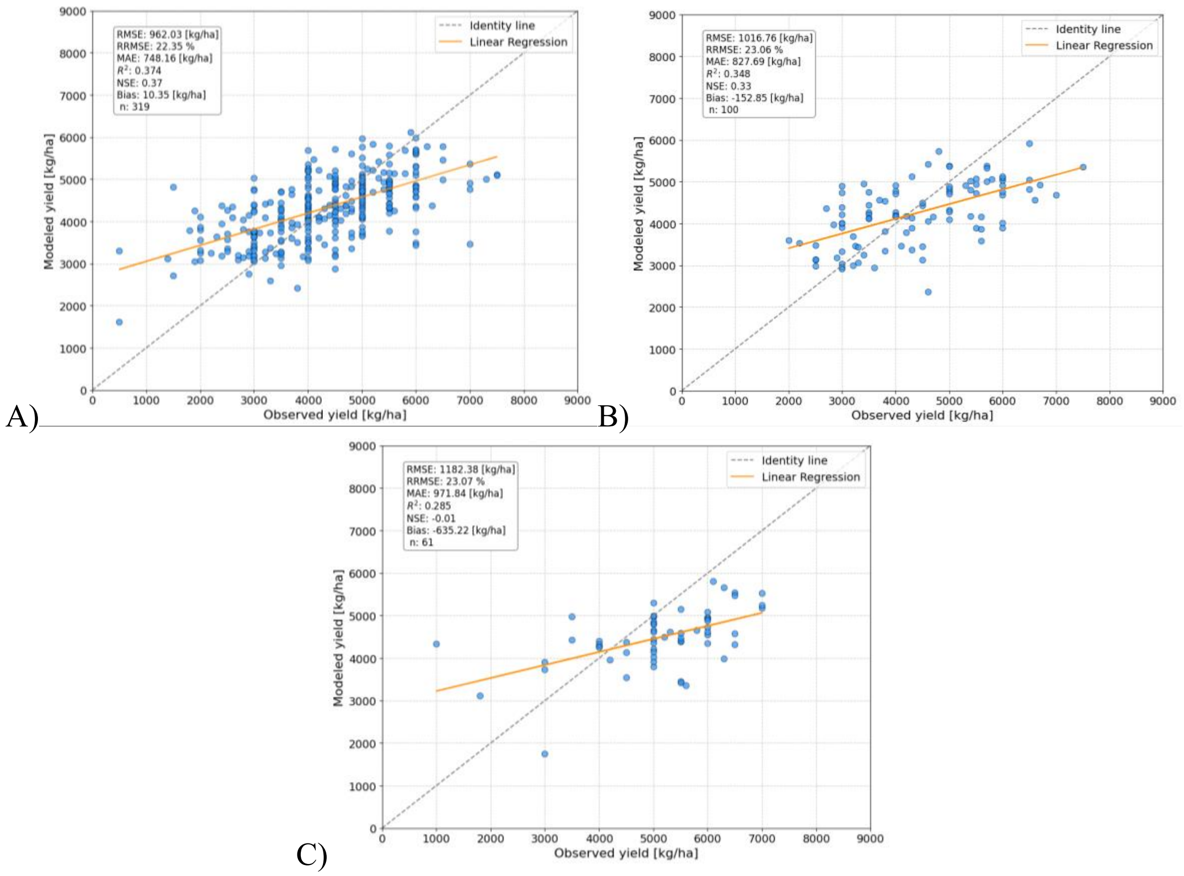


Figure 51: Comparison of modeled yield using the Sen4Stat toolbox to the observed yield for A) eye-estimated, B) non-eye-estimated, and C) irrigated parcels. Only the validation dataset was used.

6. Discussion

The aim of this study was to apply the SAFYE-CO2 model, previously tested only on wheat and sunflower in France, to barley production in Spain, in order to evaluate the potential of this process-based model in a new but agronomically comparable context. Particular emphasis was placed on the model's scalability by assessing its performance at both parcel and provincial levels, its capacity to estimate yield using a restricted LAI dataset, and its integration within the ACEO processing chain. This work contributes to ongoing efforts in advancing yield estimation methodologies from the parcel to the global scale. The following section discusses the results in relation to the research objectives and situates them within the broader context of current developments in yield estimation that is grouped into two dominant modeling paradigms: statistical and process-based modeling.

6.1 Sensitivity and uncertainty analysis

The first research objective was to identify relevant parameters to be calibrated and their potential uncertainty. The sensitivity analysis shows that all five parameters that were identified in the literature to be calibrated by parcel, influence the LAI and/or yield. However, it also highlighted the persisting issue of equifinality, as the Bayesian Optimization process resulted in different parameter combinations for the same parcel. While this could partly be due to the Bayesian Optimization algorithm not always converging to the global minima like a grid search approach would, the analysis showed that different parameter combinations actually resulted in the same LAI curves, indicating that the problem lies in the parameter identifiability.

In particular, the parameters SLAb, PRT_La, and LUEa were identified as difficult to calibrate. As shown in the previous section, these are exactly the parameters to be most influential for yield estimation. While they were still calibrated in this study, this finding has important implications for future applications of the model.

Firstly, a large part of the uncertainty can be attributed to the parameter LUEa. Since this is considered as one of the most important parameters (Duchemin et al., 2008), better estimation of it, e.g., by using other methods and indices, such as the chlorophyll content or albedo (Dong et al., 2017; H. Wang et al., 2010), could be considered.

Secondly, the ranges of the parameters SLAb and PRT_La are not yet well understood, as the sensitivity analysis showed that certain parameter combinations may produce unrealistic values. While in most cases the unrealistic combinations are avoided through LAI assimilation, this is an indication that the interaction between the parameters is not yet fully understood and represented in the modeling process. By better defining and restricting the parameter ranges, not only could the accuracy and realism of the model be improved, but the calibration process could be more efficient. Alternatively, setting one parameter helped to reduce the uncertainty and complexity (see Section 5.3.1).

On the other hand, in certain cases, additional complexity may be justified to better capture key physiological or external processes. For instance, in SAFYE-CO2, yield is modeled simply as a fixed proportion of DAM. Yet, grain development is known to be highly sensitive to meteorological conditions during anthesis and grain-filling stages (Liu et al., 2016; Yiğit &

Chmielewski, 2024). This suggests that incorporating these dynamics could improve model realism, albeit at the cost of added complexity. Similarly, the current representation of senescence relies on just two parameters (the onset and rate of senescence) and has limited influence on yield outcomes. However, late-season weather events, such as heavy rainfall, can delay harvest operations or increase disease risk, thereby affecting actual yield (van der Velde et al., 2012). These factors, while external to plant growth per se, can be relevant and are accounted for in more complex models, such as APSIM (Holzworth et al., 2014; Keating et al., 2003).

Ultimately, a central challenge in process-based model development is striking the right balance between complexity and interpretability. While additional parameters can enhance biological realism, they also elevate the risk of overfitting and equifinality if not properly constrained.

6.2 Yield estimation at parcel level

Despite the limitations identified through the sensitivity analysis, yield estimation at parcel level was performed to investigate the second research objective, namely the potential of SAFYE-CO2 to estimate barley yield in Spain using LAI assimilation.

Based on the low RMSE between modeled and remotely-sensed LAI in Figures 29, 31, and 33, as well as the good correlation coefficient between LAI and yield, the LAI assimilation can be considered successful. In addition, several noteworthy insights emerged from the analysis, highlighted below.

Despite the good correlation coefficient between LAI and yield, a saturation effect is visible. Similarly, Duveiller et al. (2011) observed a saturation effect when using an inverse modeling approach to estimate the GAI from remote sensing observations. They attributed this underestimation of GAI at high LAI values to shortcomings in the model representation of the canopy, equifinality issues in inverse modeling, and signal saturation. To address these issues, they proposed applying a linearization, as a linear relationship is expected between ground-measured and modelled GAI/LAI. This approach could be considered to mitigate the observed yield saturation at higher LAI values, particularly since the model remains limited in its representation of yield development and, as shown above, is affected by equifinality. Therefore, exploring a linearization method to reduce yield saturation at high LAI values warrants further investigation.

A limitation of the applied calibration process lies in the detection of emergence and harvest dates, as well as the preprocessing of the LAI data, which can introduce errors in the modeling process. In particular, for a few cases, the harvest date was set too early, ending plant growth prematurely and significantly underestimating the yield. This could be avoided if reported harvest dates were available or by setting a generic late harvest date to capture the end of the senescence for all parcels. Furthermore, some parcels exhibited two peeks during the season, making it difficult to identify the barley growth season. A better cleaning by excluding parcels with two LAI peaks could be applied. Nonetheless, for the majority of the parcels a good agreement between the observed and modeled LAI curve was achieved.

Moreover, the calibration of the HI based on calibration parcels identified the HI between 0.53 and 0.57 for SENa 1, and 0.45 and 0.46 for SENa 2. These values are in line with HIs identified in previous literature, for example Plaza-Bonilla et al. (2017) identified the HI for rainfed barley in Spain between 0.39 - 0.53. The significantly lower calibrated HI for SENa 2 can be explained by the higher SENa values found in this approach, which support prolonged biomass accumulation and, consequently, higher yield production. By calibrating the HI to a lower value, the increased biomass production is compensated for. Additionally, it was noted that both of the SENa detection methods occasionally set the SENa too early or too late, which affected a successful LAI assimilation. Since this affects the calibration of the remaining parameters, it could be considered to include SENa in the calibration process.

Nevertheless, the SENa 1 method consistently showed slightly better performance than SENa 2 and achieved comparable accuracies to previous SAFY-model applications. As reviewed, RMSE values in the literature (at parcel as well as provincial level) range from 0.36 – 1.39 t/ha, with RRMSE from 4.6 – 23.3 %. The best RMSE values in this study (1.1 – 1.5 t/ha, using SENa 1) are at the upper end of this range, yet with a slightly higher RRMSE ranging from 27.4 – 31.8%. The most comparable study by Pique et al. (2020a), achieved an RMSE = 1.02 t/ha for 16 wheat fields. However, a direct comparison is limited due to differences in sample size and data quality. This study relies on observational data from over 1000 non-controlled fields, in contrast to smaller, controlled experiments. This is particularly evident in model fit: SAFYE-CO2 showed low agreement with observed yield trends, with R^2 ranging from 0.11 to 0.22, compared to 0.35–0.96 in other studies and $R^2 = 0.78$ in Pique et al. (2020a).

While model performance and parameter calibration may partly explain the low R^2, the relatively strong correlation between LAI and yield suggests that the quality of the observational data is the primary limiting factor. Observed yield values often remain constant across a wide range of LAI, indicating either that the model misses explanatory variables outside the LAI, climate, or soil parameters which cause this distribution in values, or more likely that the observed yield is highly biased.

The impact of the particular spread of the observational data is also reflected in the model's bias, which, at parcel level, is low in comparison to the RMSE. It appears that there is no systematic error in the model, meaning the main issue is not a systematic over- or underestimation but rather random variation and poor prediction accuracy at the individual parcel level with the model frequently failing to accurately predict the yield.

To answer the second research objective, the HI was successfully calibrated as it is in line with previous literature, despite challenges due to low correlation between the observed and modeled yield. While SAFYE-CO2 can predict the yield for rainfed parcels with an error of approximately 1.2 t/ha, the accuracy is insufficient to provide reliable results at small scales. In addition, the model performs poorly in capturing the trend and variability in the observed yield, likely due in part to the quality and uncertainty of parcel-level yield observations.

6.3 Yield estimation at provincial level

Aggregating observed and modeled yield at provincial level has significantly improved the model's accuracy. The modeled yield varies by less than 10% of the mean observed yield and improved the R^2 to 0.36-0.65. Similar findings were observed for example, by Deines et al. (2021). Applying the SCYM at parcel level, they achieved an R^2 of 0.4 at pixel level which improved to 0.45 at field level and 0.69 at county level. Similarly, Njogu (2024) using statistical models on wheat in the same region observed improvements from RMSE = 940-969 kg/ha, RRMSE = 24%, and R^2 = 0.56-0.59 to RMSE = 315-365 kg/ha, RRMSE = 7.8 - 9%, and R^2 = 0.7-0.78

This is the result of averaging out errors. As discussed in the previous section, the model randomly over- and underestimates yield. By aggregating the values per province, the errors in both directions balance each other out.

To conclude and to answer the third research objective, this study supports previous findings that aggregating results at provincial level increases performance. For governmental institutes that require yield estimates at provincial, regional or national levels, process-based models could therefore supplement sampling survey methods. However, while the performance seems to have improved, it also highlights how well aggregated results can hide large discrepancies at smaller scales. If average provincial yield estimates hide large within-province variability, the provincial average offers limited value to farmers, as actual yields may vary by several tons per hectare, emphasizing the desire for higher spatial resolution estimates.

6.4 Yield estimation with restricted LAI dataset

The goal of the fourth research objective was to identify whether the model would be able to extract the necessary parameters correctly even from a shorter time period, i.e., up to approximately the point of maximum LAI, by using a restricted dataset of one to two months pre-harvest. This is valuable information when considering that most yield estimates are desired pre-harvest to allow planning of storage capacity, prices, and trading (van der Velde et al., 2019).

Even with a restricted LAI dataset, the model performed only slightly worse. There are multiple explanations for this. Firstly, restricting the LAI to one to two months pre-harvest still allows assimilation of the LAI up to maximum LAI for most parcels. Four out of the five calibrated parameters are related to the LAI curve until the maximum LAI, i.e., the growth of the leaves. Hence, these parameters are likely well captured. Secondly, the sensitivity analysis revealed that the most influential parameter for yield, is PRT_Le. PRT_Le partly determines until when biomass is accumulated, hence, if this parameter can be well captured from the left side of the LAI curve, then a good yield estimation is possible even when the senescence parameter SENb is not properly calibrated. Lastly, the DGM is a proportion of the DAM. If the relevant parameters for the DAM production are extracted from the curve until max LAI, the DGM should be accurately estimated as well. Biologically, this is explained by the fact that the grain filling takes place until the maximum LAI is reached (Meier, 2001), hence, most of the yield development takes up to this point.

While this is an encouraging result and in line with previous research as outlined in Section 2.6.1 (Sisheber et al., 2024), for the actual implementation of SAFYE-CO2 in forecasting mode, additional factors need to be considered. For example, the SENa and harvest parameters were still determined from the full LAI curve, which would not be available pre-harvest.

Additionally, the full meteorological time series was used. In forecasting mode, weather predictions for one to two months are likely not as accurate. Hence, the accuracy of a proper yield forecast would also depend on the accuracy of the weather forecast.

Nonetheless, this exercise provided valuable insights into the relevance of the calibrated parameters with the yield being more dependent on the calibration of growth parameters than on that of senescence parameters. These insights can be used for further model adjustments as explained in more detail in the following section.

6.5 Application of ACEO for national operability

The goal of the fifth research objective was to investigate the scaling of SAFYE-CO2 using the ACEO processing chain. The concept behind ACEO is to use a LUT calibrated at pixel level to efficiently generate yield estimates at high spatial resolution. This should enable national yield statistics at much finer detail to account for spatial diversity.

However, ACEO performs worse than SAFYE-CO2 at parcel level, which is shown in insignificant R^2 values. The results show no discrimination between parcels of higher and lower LAI and the total range of the yield is smaller than the observed range: ACEO covers a range of approximately 5 t/ha, whereas the observational data covers a range of 7 t/ha. Considering that the individual calibration at parcel level of SAFYE-CO2 lead to better results, the poor performance seems more connected to the implementation of the ACEO processing chain. For example, instead of using interpolated and smoothed LAI curves, ACEO only uses the observed LAI values. This results in fewer values used for the identification of the suitable parameter set and likely identifies different parameter sets that produce very different yields, which eventually can average each other out. This could have contributed to the weak association between modeled yield and LAI, and observed yield. Furthermore, the sampling strategy of the parameter distribution can be investigated, as ACEO currently samples from a normal distribution. A comparison to the parameter distribution when applying SAFYE-CO2 individually could provide useful insights.

Compared to the SCYM approach, ACEO currently shows lower performance in yield estimation. The first implementation by Lobell et al. (2015) on 17,000 maize fields and 11,000 soybean fields in the US achieved an average R^2 of 0.35 for maize and 0.32 for soybean at field level. Deines et al. (2021) applied the SCYM across 10 years for corn fields in the US and achieved an R^2 of 0.45 at field level. However, this needs to take into account that the study area consisted of large, homogeneous, irrigated and highly productive fields. This leads to less variability between fields, making crop growth simulation less complex. A better comparison would be the application to small-holder farms in Uganda, where Lobell et al. (2020) achieve an R^2 between 0.28 and 0.58, depending on the observational ground-truth data used for comparison.

This suggests that further refinement and methodological development are necessary to enhance ACEO's scalability and operational utility. One key limitation is that ACEO requires regenerating the LUT for each year to account for changes in weather conditions, which limits its efficiency. In contrast, SCYM relies on a statistical model trained on outputs from multiple years of process-based simulations, allowing it to generalize across seasons without needing yearly recalibration. Integrating a similar approach into ACEO could be a valuable step toward making it more robust and scalable for operational yield forecasting.

In addition to improving the ACEO performance, certain conditions for successful national to global crop yield estimation need to be considered.

It is important to note that processing all barley parcels within a single Sentinel-2 tile, which covered only part of the study area, took nearly four days. Therefore, to facilitate broader application, efforts should be made to improve the user-friendliness of the processing chain or to ensure that users have access to adequate computational resources.

Moreover, to apply the model, good knowledge of the crop land area is necessary. This is a challenge not unique to process-based models (Fritz et al., 2019), but very relevant, since the location and area needs to be known to estimate yield and production. While Spain has a good coverage of its crop land area, for low-income countries the infrastructure to record crop type coverage is poor (Deines et al., 2021; Fritz et al., 2019). But even at European level, high-resolution seasonal cropland maps that display area by crop type are only recently being developed with the advances of high-resolution satellite missions (Fan et al., 2021).

Furthermore, this study profited from available high-resolution LAI data to perform the LAI assimilation. However, in areas with persistent cloud cover, the number of usable satellite acquisitions can be very low. This remains an issue when using optical remote sensing data over short agricultural periods (Defourny et al., 2019).

6.6 Comparison of Process-Based Models to Statistical Models

To address the final research objective, SAFYE-CO2 results are compared to those obtained using the Sen4Stat statistical model. As shown in Section 5.9, Sen4Stat achieved a slightly better performance, with lower RMSE and higher R^2. This may be due to the model being directly trained on the observational data, allowing it to better fit the observed variability. Furthermore, the statistical model required only a small calibration dataset and was rapidly operationalized. In contrast, process-based models require careful selection of appropriate parameters and their calibration. As shown in Section 6.1, striking the right balance between complexity and interpretability remains a central challenge in process-based model development. This raises a broader question in yield modeling: whether the pursuit of mechanistic detail in process-based models offers a tangible advantage over simpler data-driven approaches. As discussed in the literature review, proponents of process-based approaches emphasize their ability to represent underlying biophysical processes, enhancing transferability across time and space. The following sections critically examine these claims in comparison to statistical models.

The first advantage of process-based models in comparison to statistical models usually highlighted is the use of less training data. However, even if process-based models require less observational data than statistical models, observations are still important to validate the model. As was shown in this study, the quality of the yield observations is highly important. Given the biases in the observations, their quality remains questionable, making evaluation of the model performance difficult.

The second advantage is the transferability to different locations, growing conditions and plant types, due to the consideration of soil-plant-atmosphere-management in process-based models. Nevertheless, this study showed that a significant amount of calibration is required, as well as a good understanding of the right selection and values of parameters. While most of the fixed parameter values were available for barley due to its similarity to wheat, for other crops, or varieties with characteristics differing by ecoregion, these parameters or the ranges of the calibrated parameters would need to be determined. Identification of such parameters can be time-consuming and require in-field experiments which can limit fast adaptation of the model to different study areas and use cases, where statistical models might be trained faster, if sufficiently high-quality training data are available (Lobell & Asseng, 2017).

The third advantage usually attributed to process-based models is their applicability under climate change scenarios. It is assumed that process-based models can account for extreme weather conditions as they are not restricted by training data, which is the case for statistical models (Lobell & Burke, 2010). For example, in SAFYE-CO2, the temperature stress is modeled using an exponential equation, which decreases productivity when air temperature deviates from the optimal temperature. However, the year of this study was an exceptionally high performing season, hence, a comparison to a year with for example heat waves would be interesting. Many extremes also exaggerate their effect depending on the occurrence during the phenological stages and the sequence in which they occur, e.g., multiple hot and dry days. Including such factors can make the model even more complex, which opens the question whether including such factors is feasible to implement in process-based models, or whether it is faster to develop statistical models. For example, in a first attempt to investigate the ability of machine learning models to predict yield under climate scenarios of +2°C, Lobell & Burke (2010) find that statistical models, trained with simulations from a process-based model, relatively well reproduce the effect of climate change on yield. Similarly, Lobell & Asseng (2017) compare the sensitivity of process-based models and statistical models, and find that both show similar impacts on yield through temperature, but not for precipitation, where process-based models are more sensitive. Overall, when modeling yield under climate change scenarios, both process-based and statistical models need to be improved. Lobell & Asseng (2017) propose that one way is to combine their strengths, using process-based models to better understand the effects of meteorological events and identify important variables that can be used as inputs for the training of statistical models.

Hence, no universal answer as to which model should be preferred exists, yet it remains a fundamental debate that encourages new directions for the future of model development (Baker et al., 2018). For example, process-based models can be used not only as predictive tools but also as interpretative frameworks. They are valuable for understanding the processes, parameters, and phenological stages that drive crop development. This mechanistic insight can complement and even inform statistical models, particularly in hybrid modeling frameworks that benefit from physically meaningful parameter inputs (Lobell & Asseng, 2017). An example is the Sen4Stat model, which already incorporates features extracted from the SAFY model into its processing chain. Since the SAFYE-CO2 model has significantly developed from its initial version, it would be worthwhile to investigate whether the enhanced level of detail of SAFYE-CO2 could be integrated into, and potentially improve, the Sen4Stat model.

7. Conclusion

This study aimed to advance research in the field of high-resolution crop yield estimation across diverse landscapes with small parcel sizes, with the goal of enabling scalable applications from farm to national and global levels. To capture spatial variability at high resolution for small parcel sizes, Sentinel-2-derived LAI data were used to provide parcel-level input. Different modeling techniques were tested, with a focus on the process-based model SAFYE-CO2.

Parcel-wise application of the SAFYE-CO2 model produces results comparable to previous studies, but are situated on the lower-performing end. The analysis highlights several opportunities to improve model performance. Firstly, adjusting the relationship between modeled yield and observed LAI could lead to a better representation of high yield values. Alternatively, a more realistic representation of the grain development process could lead to more accurate and robust results. Secondly, improving the identifiability and prior ranges of the calibrated parameters is needed to address equifinality issues. In particular, more attention to the LUEa parameter should be paid. Alternative methods for estimating LUEa could be explored. Thirdly, assessing the potential advantages of calibrating the SENa parameter instead of extracting it from the LAI curve, as well as setting a generic late harvest date, could improve model performance. Lastly, the study underlines the critical need for more reliable, fine-scale ground truth data representative at parcel instead of provincial level to evaluate and validate yield estimates at high spatial resolutions. For example, sampling surveys based on full crop cuts could offer more accurate benchmarks.

A notable result is the model's comparable performance when using a restricted LAI dataset. This suggests that calibration of the current selection of parameters may be feasible pre-harvest, opening promising avenues for yield forecasting using the SAFYE-CO2 model. Nonetheless, apart from improving the model accuracy in general, yield forecasts would also require incorporating meteorological forecasts and evaluating their influence on yield forecasting accuracies.

To investigate the possibility of scaling the process-based model to provide high-spatial resolution yield estimates for large areas, the AgriCarbon-EO processing chain, which integrates the SAFYE-CO2 model, was applied. While it provides pixel-wise yield estimates, two main challenges emerged that require further development of the processing chain. From a technical standpoint, the implementation remains processing-intensive, which requires users to have sufficient computational resources to be able to implement the processing chain at regional to national scales. It also depends on the availability of accurate geolocation data to delineate individual fields. However, not all countries have the necessary infrastructure or systems in place to collect and maintain such data, as is the case in Spain. Generating detailed parcel shapefiles often requires significant effort and technical expertise that may be limited depending on the country. In such contexts, recent advances in remote sensing–based methods for identifying cropland areas and crop types offer a promising alternative. From a performance perspective, the ACEO performed poorly. This suggests the need for more development to achieve results similar to when implementing SAFYE-CO2 parcel-wise. Improved use of LAI data through cleaning and interpolation, and improved understanding of the parameter distributions, could help address this issue.

Lastly, the study highlights that although statistical models are easier to implement and perform well, their over-reliance on training data limits their ability to be scaled across regions or to adapt to novel conditions. Instead, a hybrid modeling approach, such as the ACEO or SCYM methods, appears more promising. These methods combine the computational efficiency of statistical models with the physiological and phenological understanding of process-based models. Advancing and improving such methods likely continues to be the best way forward, to enable yield estimates at resolutions much higher than currently available at large scales.

In summary, while challenges remain in applying process-based models at field scale and over large areas, this study contributes to the ongoing discourse on their role in agricultural monitoring. It encourages future research to refine model design, integrate more reliable field data, and explore hybrid methods that leverage the strengths of both statistical and process-based modeling approaches for improved yield estimation across scales. Ultimately, the development of robust, scalable, and interpretable yield estimation tools serves a broad range of stakeholders, from farmers for optimizing inputs to policymakers for evaluating food security risks, highlighting the need for both scientific rigor and operational feasibility in future model development.

References

Abd-Elmabod, S. K., Muñoz-Rojas, M., Jordán, A., Anaya-Romero, M., Phillips, J. D., Jones, L., Zhang, Z., Pereira, P., Fleskens, L., van der Ploeg, M., & de la Rosa, D. (2020). Climate change impacts on agricultural suitability and yield reduction in a Mediterranean region. *Geoderma*, *374*, 114453. https://doi.org/10.1016/j.geoderma.2020.114453

Al Bitar, A., Wijmer, T., & Ceschia, E. (2023). *SAFYE-CO2 Python code manual* (Issue 2.0.5).

Baker, R. E., Peña, J. M., Jayamohan, J., & Jérusalem, A. (2018). Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biology Letters*, *14*(5), 1–4. https://doi.org/10.1098/rsbl.2017.0660

Ballabio, C., Panagos, P., & Monatanarella, L. (2016). Mapping topsoil physical properties at European scale using the LUCAS database. *Geoderma*, *261*, 110–123. https://doi.org/10.1016/j.geoderma.2015.07.006

Baret, F., & Guyot, G. (1991). Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment*, *35*(2–3), 161–173. https://doi.org/10.1016/0034-4257(91)90009-U

Baruth, B., Bassu, S., Bussay, A., Ceglar, A., Cerrani, I., Chemin, Y., De Palma, P., Fumagalli, D., Lecerf, R., Manfron, G., Nisini Scacchiafichi, L., Panarello, L., Ronchetti, G., Seguini, L., Toreti, A., Van Den Berg, M., Van Der Velde, M., Zajac, Z., Zucchini, A., … Niemeyer, S. (2020). *JRC MARS Bulletin - Vol. 28 No 5, Crop monitoring in Europe - May 2020* (Issue KJ-AW-20-005-EN-N (online),KJ-AW-20-005-EN-C (print)). Publications Office of the European Union.

Basso, B., & Liu, L. (2019). Seasonal crop yield forecast: Methods, applications, and accuracies. *Advances in Agronomy*, *154*, 201–255. https://doi.org/10.1016/bs.agron.2018.11.002

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

Bauer-Marschallinger, B., & Paulik, C. (2019). Copernicus Global Land Operations “Vegetation and Energy”, PRODUCT USER MANUAL, SURFACE SOIL MOISTURE COLLECTION 1KM VERSION 1. *Copernicus Global Land Operations*, *I1.30*.

Becker-Reshef, I., Barker, B., Humber, M., Puricelli, E., Sanchez, A., Sahajpal, R., McGaughey, K., Justice, C., Baruth, B., Wu, B., Prakash, A., Abdolreza, A., & Jarvis, I. (2019). The GEOGLAM crop monitor for AMIS: Assessing crop conditions in the context of global markets. *Global Food Security*, *23*, 173–181. https://doi.org/10.1016/j.gfs.2019.04.010

Berjón, A. J., Cachorro, V. E., Zarco-Tejada, P. J., De Frutos, A. M., & Martin, P. (2003). Estimation of Leaf Area Index and chlorophyll content in barley by inversion of radiative transfer models at different growth stages. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *4*, 2189 – 2191. https://www.scopus.com/inward/record.uri?eid=2-s2.0-0242626633&partnerID=40&md5=37892736b7946796c445be025fe92756

Bindi, M., Miglietta, F., & Zipoli, G. (1992). Different methods for separating diffuse and direct components of solar radiation and their application in crop growth models. *Climate Research*, *2*(1), 47–54. http://www.jstor.org/stable/24863284

Bontemps, S., Cara, C., Deffense, N., Nicola, L., Nörgaard, B., & Udroiu, C. (2024). *ESA Sen4Stat Sentinels for Agricultural Statistics, Software User Manual* (Issue I.3). https://www.esa-sen4stat.org/wp-content/uploads/content/Sen4Stat_SUM_v1.3.pdf

Boogaard, H., Wolf, J., Supit, I., Niemeyer, S., & van Ittersum, M. (2013). A regional implementation of WOFOST for calculating yield gaps of autumn-sown wheat across the European Union. *Field Crops Research*, *143*, 130–142. https://doi.org/10.1016/j.fcr.2012.11.005

Cai, Z., Jönsson, P., Jin, H., & Eklundh, L. (2017). Performance of Smoothing Methods for Reconstructing NDVI Time-Series and Estimating Vegetation Phenology from MODIS Data. *Remote Sensing*, *9*(12), 1271. https://doi.org/10.3390/rs9121271

Cammarano, D., Ceccarelli, S., Grando, S., Romagosa, I., Benbelkacem, A., Akar, T., Al-Yassin, A., Pecchioni, N., Francia, E., & Ronga, D. (2019). The impact of climate change on barley yield in the Mediterranean basin. *European Journal of Agronomy*, *106*, 1–11. https://doi.org/10.1016/j.eja.2019.03.002

Candelieri, A. (2021). A Gentle Introduction to Bayesian Optimization. *2021 Winter Simulation Conference (WSC)*, 1–16. https://doi.org/10.1109/WSC52266.2021.9715413

Cao, J., Zhang, Z., Luo, Y., Zhang, L., Zhang, J., Li, Z., & Tao, F. (2021). Wheat yield predictions at a county and field scale with deep learning, machine learning, and google earth engine. *European Journal of Agronomy*, *123*, 126204. https://doi.org/10.1016/j.eja.2020.126204

Carletto, C., Jolliffe, D., & Banerjee, R. (2015). From Tragedy to Renaissance: Improving Agricultural Data for Better Policies. *The Journal of Development Studies*, *51*(2), 133–148. https://doi.org/10.1080/00220388.2014.968140

Ceschia, E., Al Bitar, A., Wijmer, T., Pique, G., Fieuzal, R., & Arnaud, L. (n.d.). *SAFYE CO2 Version 2.0.5*. CESBIO (Université de Toulouse, CNES, CNRS, UPS, IRD, INRAE).

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. https://doi.org/10.1080/01431161.2013.875629

Chahbi Bellakanji, A., Zribi, M., Lili-Chabaane, Z., & Mougenot, B. (2018). Forecasting of Cereal Yields in a Semi-arid Area Using the Simple Algorithm for Yield Estimation (SAFY) Agro-Meteorological Model Combined with Optical SPOT/HRV Images. *Sensors*, *18*(7), 2138. https://doi.org/10.3390/s18072138

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Chipanshi, A., Zhang, Y., Kouadio, L., Newlands, N., Davidson, A., Hill, H., Warren, R., Qian, B., Daneshfar, B., Bedard, F., & Reichert, G. (2015). Evaluation of the Integrated Canadian Crop Yield Forecaster (ICCYF) model for in-season prediction of crop yield across the Canadian agricultural landscape. *Agricultural and Forest Meteorology*, *206*, 137–150. https://doi.org/10.1016/j.agrformet.2015.03.007

Climate Change Knowledge Portal. (n.d.). *Observed Climate Data, CRU TS4.08 0.5-Degree*. https://doi.org/https://doi.org/10.57966/tw2k-9h36

de Jong, J. B. R. M. (1980). Een karakterisering van de zonnestraling in Nederland. *Technische Hogeschool Eindhoven*.

de Wit, A., Boogaard, H., Fumagalli, D., Janssen, S., Knapen, R., van Kraalingen, D., Supit, I., van der Wijngaart, R., & van Diepen, K. (2019). 25 years of the WOFOST cropping systems model. *Agricultural Systems*, *168*, 154–167. https://doi.org/10.1016/j.agsy.2018.06.018

Deffense, N., Cara, C., Nicola, L., Udroiu, C., & Bontemps, S. (2021). *ESA Sen4Stat Sentinels for Agricultural Statistics, Algorithm Theoretical Basis Document: Spectral Indices & Biophysical Indicators*.

Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Nicola, L., Rabaute, T., Savinaud, M., Udroiu, C., Valero, S., Bégué, A., Dejoux, J. F., El Harti, A., Ezzahar, J., Kussul, N., Labbassi, K., … Koetz, B. (2019). Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world. *Remote Sensing of Environment*, *221*(December 2018), 551–568. https://doi.org/10.1016/j.rse.2018.11.007

Deines, J. M., Patel, R., Liang, S.-Z., Dado, W., & Lobell, D. B. (2021). A million kernels of truth: Insights into scalable satellite maize yield mapping and yield gap analysis from an extensive ground dataset in the US Corn Belt. *Remote Sensing of Environment*, *253*, 112174. https://doi.org/10.1016/j.rse.2020.112174

Dhillon, M. S., Dahms, T., Kuebert-Flock, C., Borg, E., Conrad, C., & Ullmann, T. (2020). Modelling Crop Biomass from Synthetic Remote Sensing Time Series: Example for the DEMMIN Test Site, Germany. *Remote Sensing*, *12*(11), 1819. https://doi.org/10.3390/rs12111819

Di Bucchianico, A. (2007). Coefficient of Determination (R 2). In *Encyclopedia of Statistics in Quality and Reliability*. Wiley. https://doi.org/10.1002/9780470061572.eqr173

Diskin, P. (1999). Agricultural Productivity Indicators Measurement Guide, Arlington, Va: Food Security and Nutrition Monitoring (IMPACT) Project, ISTI. *Methods*, *January*. http://pdf.usaid.gov/pdf_docs/Pnacg169.pdf

Dong, T., Liu, J., Qian, B., Jing, Q., Croft, H., Chen, J., Wang, J., Huffman, T., Shang, J., & Chen, P. (2017). Deriving Maximum Light Use Efficiency From Crop Growth Model and Satellite Data to Improve Crop Biomass Estimation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(1), 104–117. https://doi.org/10.1109/JSTARS.2016.2605303

Duan, S.-B., Li, Z.-L., Wu, H., Tang, B.-H., Ma, L., Zhao, E., & Li, C. (2014). Inversion of the PROSAIL model to estimate leaf area index of maize, potato, and sunflower fields from unmanned aerial vehicle hyperspectral data. *International Journal of Applied Earth Observation and Geoinformation*, *26*, 12–20. https://doi.org/10.1016/j.jag.2013.05.007

Duchemin, B., Fieuzal, R., Rivera, M., Ezzahar, J., Jarlan, L., Rodriguez, J., Hagolle, O., & Watts, C. (2015). Impact of Sowing Date on Yield and Water Use Efficiency of Wheat

Analyzed through Spatial Modeling and FORMOSAT-2 Images. *Remote Sensing*, *7*(5), 5951–5979. https://doi.org/10.3390/rs70505951

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. https://doi.org/10.1016/j.envsoft.2007.10.003

Duveiller, G., & Defourny, P. (2010). A conceptual framework to define the spatial resolution requirements for agricultural monitoring using remote sensing. *Remote Sensing of Environment*, *114*(11), 2637–2650. https://doi.org/10.1016/j.rse.2010.06.001

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. https://doi.org/10.1016/j.rse.2010.11.016

ECMWF. (2024). *Conversion table for accumulated variables (total precipitation/fluxes)*. https://confluence.ecmwf.int/pages/viewpage.action?pageId=197702790

European Parliament Council of the European Union. (2021). *Regulation (EU) 2018/1091 of the European Parliament and of the Council of 18 July 2018 on integrated farm statistics and repealing Regulations (EC) No 1166/2008 and (EU) No 1337/2011 (Text with EEA relevance.)*.

European Union's Copernicus Land Monitoring Service information. (2025). *Surface Soil Moisture 2014-present (raster 1 km), Europe, daily – version 1 [Data set]*. https://land.copernicus.eu/en/products/soil-moisture/daily-surface-soil-moisture-v1.0#download

Eurostat. (2025). *Crop production in EU standard humidity [Data set]*. https://doi.org/https://doi.org/10.2908/APRO_CPSH1

Fan, J., Defourny, P., Zhang, X., Dong, Q., Wang, L., Qin, Z., De Vroey, M., & Zhao, C. (2021). Crop Mapping with Combined Use of European and Chinese Satellite Data. *Remote Sensing*, *13*(22), 4641. https://doi.org/10.3390/rs13224641

FAO. (2017). The future of food and agriculture: trends and challenges. In *The future of food and agriculture: trends and challenges*. www.fao.org/publications%0Ahttp://www.fao.org/3/a-i6583e.pdf%0Ahttp://siteresources.worldbank.org/INTARD/825826-1111044795683/20424536/Ag_ed_Africa.pdf%0Awww.fao.org/cfs%0Ahttp://www.jstor.org/stable/4356839%0Ahttps://ediss.uni-goettingen.de/bitstream/han

FAO, IFAD, UNICEF, WFP, & WHO. (2021). *The State of Food Security and Nutrition in the World 2021. Transforming food systems for food security, improved nutrition and affordable healthy diets for all.* https://doi.org/10.4060/cb4474en

Fernández, J., Frías, M. D., Cabos, W. D., Cofiño, A. S., Domínguez, M., Fita, L., Gaertner, M. A., García-Díez, M., Gutiérrez, J. M., Jiménez-Guerrero, P., Liguori, G., Montávez, J. P., Romera, R., & Sánchez, E. (2019). Consistency of climate change projections from multiple global and regional model intercomparison projects. *Climate Dynamics*, *52*(1–2), 1139–1156. https://doi.org/10.1007/s00382-018-4181-8

Fritz, S., See, L., Bayas, J. C. L., Waldner, F., Jacques, D., Becker-Reshef, I., Whitcraft, A., Baruth, B., Bonifacio, R., Crutchfield, J., Rembold, F., Rojas, O., Schucknecht, A., Van der Velde, M., Verdin, J., Wu, B., Yan, N., You, L., Gilliams, S., … McCallum, I. (2019). A comparison of global agricultural monitoring systems and current gaps. *Agricultural Systems*, *168*, 258–272. https://doi.org/10.1016/j.agsy.2018.05.010

Garrachón-Gómez, E., García, I., García-Rodríguez, A., García-Rodríguez, S., & Alonso-Tristán, C. (2024). Monthly intercepted photosynthetically active radiation estimation based on the Beer-Lambert's law across the cereal crops of Castilla y León (Spain). *Computers and Electronics in Agriculture*, *216*, 108523. https://doi.org/10.1016/j.compag.2023.108523

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Holzworth, D. P., Huth, N. I., DeVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., Murphy, C., Moore, A. D., Brown, H., Whish, J. P. M., Verrall, S., Fainges, J., Bell, L. W., Peake, A. S., Poulton, P. L., Hochman, Z., … Keating, B. A. (2014). APSIM – Evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software*, *62*, 327–350. https://doi.org/10.1016/j.envsoft.2014.07.009

IPCC. (2018). Summary for Policymakers. In V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. B. R. Matthews, Y. Chen, X. Zhou, M. I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, & T. Waterfield (Eds.), *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change,*.

Keating, B. ., Carberry, P. ., Hammer, G. ., Probert, M. ., Robertson, M. ., Holzworth, D., Huth, N. ., Hargreaves, J. N. ., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J. ., Silburn, M., Wang, E., Brown, S., Bristow, K. ., Asseng, S., … Smith, C. . (2003). An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, *18*(3–4), 267–288. https://doi.org/10.1016/S1161-0301(02)00108-9

Li, S., Huang, J., Xiao, G., Huang, H., Sun, Z., & Li, X. (2024). Improved Winter Wheat Yield Estimation by Combining Remote Sensing Data, Machine Learning, and Phenological Metrics. *Remote Sensing*, *16*(17), 3217. https://doi.org/10.3390/rs16173217

Liashchynskyi, P., & Liashchynskyi, P. (2019). *Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS*. http://arxiv.org/abs/1912.06059

Liu, B., Asseng, S., Liu, L., Tang, L., Cao, W., & Zhu, Y. (2016). Testing the responses of four wheat crop models to heat stress at anthesis and grain filling. *Global Change Biology*, *22*(5), 1890–1903. https://doi.org/10.1111/gcb.13212

Lobell, D. (2010). *Crop Responses to Climate: Time-Series Models* (pp. 85–98). https://doi.org/10.1007/978-90-481-2953-9_5

Lobell, D., & Asseng, S. (2017). Comparing estimates of climate change impacts from process-based and statistical crop models. *Environmental Research Letters*, *12*(1), 15001. https://doi.org/10.1088/1748-9326/aa518a

Lobell, D., Azzari, G., Burke, M., Gourlay, S., Jin, Z., Kilic, T., & Murray, S. (2020). Eyes in the Sky, Boots on the Ground: Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis. *American Journal of Agricultural Economics*, *102*(1), 202–219. https://doi.org/10.1093/ajae/aaz051

Lobell, D. B., Thau, D., Seifert, C., Engle, E., & Little, B. (2015). A scalable satellite-based crop yield mapper. *Remote Sensing of Environment*, *164*, 324–333. https://doi.org/10.1016/j.rse.2015.04.021

Lobell, D., & Burke, M. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology*, *150*(11), 1443–1452. https://doi.org/10.1016/j.agrformet.2010.07.008

López-Lozano, R., Duveiller, G., Seguini, L., Meroni, M., García-Condado, S., Hooker, J., Leo, O., & Baruth, B. (2015). Towards regional grain yield forecasting with 1km-resolution EO biophysical products: Strengths and limitations at pan-European level. *Agricultural and Forest Meteorology*, *206*, 12–32. https://doi.org/10.1016/j.agrformet.2015.02.021

Luo, L., Sun, S., Xue, J., Gao, Z., Zhao, J., Yin, Y., Gao, F., & Luan, X. (2023). Crop yield estimation based on assimilation of crop models and remote sensing data: A systematic evaluation. *Agricultural Systems*, *210*, 103711. https://doi.org/10.1016/j.agsy.2023.103711

Ma, C., Liu, M., Ding, F., Li, C., Cui, Y., Chen, W., & Wang, Y. (2022). Wheat growth monitoring and yield estimation based on remote sensing data assimilation into the SAFY crop growth model. *Scientific Reports*, *12*(1), 5473. https://doi.org/10.1038/s41598-022-09535-9

Martínez-Moreno, F., Solís, I., & Igartua, E. (2024). Barley History and Breeding in Spain. *Agriculture*, *14*(10), 1674. https://doi.org/10.3390/agriculture14101674

Meier, U. (2001). BBCH-Monograph Growth stages of mono- and dicotyledonous plants. In *Federal Biological Research Centre for Agriculture and Forestry*. https://www.masaf.gov.it/flex/AppData/WebLive/Agrometeo/MIEPFY800/BBCHengl2001.pdf

Mendez-Morales, M., & Calvo-Valverde, L. A. (2019). Comparison of global and local optimization methods for the calibration and sensitivity analysis of a conceptual hydrological model. *Revista Tecnología En Marcha*. https://doi.org/10.18845/tm.v32i3.4477

Ministerio de Agricultura Pesca y Alimentación. (2020). *Encuesta sobre Superficies y Rendimientos de Cultivos*.

Monteith, J. L. (1977). Climate and the efficiency of crop production in Britain. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *281*(980), 277–294. https://doi.org/10.1098/rstb.1977.0140

Murphy, J., Casley, D. J., & Curry, J. J. (1991). Farmers' estimations as a source of production data: methodological guidelines for cereals in Africa. In *World Bank Technical Paper, Africa Technical Department Series* (Vol. 132, Issue 132).

NASA. (n.d.). *MODIS Leaf Area Index/FPAR*.

Njogu, A. M. (2024). Developing a Crop Yield Estimation Method from Remotely Sensed Metrics Using Artificial Intelligence; Case Study: Spain and Western Kenya. *Faculté Des Bioingénieurs, Université Catholique de Louvain*. http://hdl.handle.net/2078.1/thesis:49081

Organisation for Economic Co-Operation and Development. (n.d.). *Crop Production*. https://www.oecd.org/en/data/indicators/crop-production.html

Panagos, P., Van Liedekerke, M., Borrelli, P., Köninger, J., Ballabio, C., Orgiazzi, A., Lugato, E., Liakos, L., Hervas, J., Jones, A., & Montanarella, L. (2022). European Soil Data Centre 2.0: Soil data and knowledge in support of the EU policies. *European Journal of Soil Science*, *73*(6). https://doi.org/10.1111/ejss.13315

Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., & Athanasiadis, I. N. (2021). Machine learning for large-scale crop yield forecasting. *Agricultural Systems*, *187*, 103016. https://doi.org/10.1016/j.agsy.2020.103016

Paudel, D., Boogaard, H., de Wit, A., van der Velde, M., Claverie, M., Nisini, L., Janssen, S., Osinga, S., & Athanasiadis, I. N. (2022). Machine learning for regional crop yield forecasting in Europe. *Field Crops Research*, *276*, 108377. https://doi.org/10.1016/j.fcr.2021.108377

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

Piferrer, S. J., Losa, S. M., & Pérez, J. J. L. (2021). *Calendario de siembra,recolección y comercialización 2014-2016*. https://www.mapa.gob.es/es/estadistica/temas/estadisticas-agrarias/01-calendariosiembra-nuevo-sencilla-1_tcm30-514260.pdf

Pique, G., Fieuzal, R., Al Bitar, A., Veloso, A., Tallec, T., Brut, A., Ferlicoq, M., Zawilski, B., Dejoux, J.-F., Gibrin, H., & Ceschia, E. (2020a). Estimation of daily CO2 fluxes and of the components of the carbon budget for winter wheat by the assimilation of Sentinel 2-like remote sensing data into a crop model. *Geoderma*, *376*, 114428. https://doi.org/10.1016/j.geoderma.2020.114428

Pique, G., Fieuzal, R., Debaeke, P., Al Bitar, A., Tallec, T., & Ceschia, E. (2020b). Combining High-Resolution Remote Sensing Products with a Crop Model to Estimate Carbon and Water Budget Components: Application to Sunflower. *Remote Sensing*, *12*(18), 2967. https://doi.org/10.3390/rs12182967

Plaza-Bonilla, D., Álvaro-Fuentes, J., Bareche, J., Masgoret, A., & Cantero-Martínez, C. (2017). Delayed Sowing Improved Barley Yield in a No-Till Rainfed Mediterranean Agroecosystem. *Agronomy Journal*, *109*(4), 1249–1260. https://doi.org/10.2134/agronj2016.09.0537

Rötter, R. P., Palosuo, T., Kersebaum, K. C., Angulo, C., Bindi, M., Ewert, F., Ferrise, R., Hlavinka, P., Moriondo, M., Nendel, C., Olesen, J. E., Patil, R. H., Ruget, F., Takáč, J., & Trnka, M. (2012). Simulation of spring barley yield in different climatic zones of Northern and Central Europe: A comparison of nine crop models. *Field Crops Research*, *133*, 23–36. https://doi.org/10.1016/j.fcr.2012.03.016

Sabins, F. F. J., & Ellis, J. M. (2020). *Remote Sensing Principles, Interpretation, and Applications* (4th Editio). Waveland Press.

Sapkota, T. B., Jat, M. L., Jat, R. K., Kapoor, P., & Stirling, C. (2016). Yield Estimation of Food and Non-food Crops in Smallholder Production Systems. In *Methods for Measuring Greenhouse Gas Balances and Evaluating Mitigation Options in Smallholder Agriculture* (pp. 163–174). Springer International Publishing. https://doi.org/10.1007/978-3-319-29794-1_8

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, *36*(8), 1627–1639. https://doi.org/10.1021/ac60214a047

Senent-Aparicio, J., López-Ballesteros, A., Jimeno-Sáez, P., & Pérez-Sánchez, J. (2023). Recent precipitation trends in Peninsular Spain and implications for water infrastructure design. *Journal of Hydrology: Regional Studies*, *45*, 101308. https://doi.org/10.1016/j.ejrh.2022.101308

Shields, M. D., & Zhang, J. (2016). The generalization of Latin hypercube sampling. *Reliability Engineering & System Safety*, *148*, 96–108. https://doi.org/10.1016/j.ress.2015.12.002

Silvestro, P. C., Pignatti, S., Yang, H., Yang, G., Pascucci, S., Castaldi, F., & Casa, R. (2017). Sensitivity analysis of the Aquacrop and SAFYE crop models for the assessment of water limited winter wheat yield in regional scale applications. *PLOS ONE*, *12*(11), e0187485. https://doi.org/10.1371/journal.pone.0187485

Sisheber, B., Marshall, M., Mengistu, D., & Nelson, A. (2024). The influence of temporal resolution on crop yield estimation with Earth Observation data assimilation. *Remote Sensing Applications: Society and Environment*, *36*, 101272. https://doi.org/10.1016/j.rsase.2024.101272

Song, Y., Wang, J., Shang, J., & Liao, C. (2020). Using UAV-Based SOPC Derived LAI and SAFY Model for Biomass and Yield Estimation of Winter Wheat. *Remote Sensing*, *12*(15), 2378. https://doi.org/10.3390/rs12152378

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), 426–437. https://doi.org/10.2134/agronj2008.0139s

Stern, A. J., Doraiswamy, P. C., & Hunt, E. R. (2014). Comparison of different MODIS data product collections over an agricultural area. *Remote Sensing Letters*, *5*(1), 1–9. https://doi.org/10.1080/2150704X.2013.862600

Tóth, B., Weynants, M., Nemes, A., Makó, A., Bilas, G., & Tóth, G. (2015). New generation of hydraulic pedotransfer functions for Europe. *European Journal of Soil Science*, *66*(1), 226–238. https://doi.org/10.1111/ejss.12192

United Nations. (2011). *The global social crisis: Report on the world social situation 2011*.

United Nations Development Programme. (2022). *What are the Sustainable Development Goals?* https://www.undp.org/sustainable-development-goals

United Nations General Assembly. (2015). *Resolution adopted by the General Assembly on 25 september 2015*. https://www.unfpa.org/sites/default/files/resource-pdf/Resolution_A_RES_70_1_EN.pdf

Upreti, D., McCarthy, T., O'Neill, M., Ishola, K., & Fealy, R. (2022). Application and Evaluation of a Simple Crop Modelling Framework: A Case Study for Spring Barley, Winter Wheat and Winter Oilseed Rape over Ireland. *Agronomy*, *12*(11), 2900. https://doi.org/10.3390/agronomy12112900

van der Velde, M., Biavetti, I., El-Aydam, M., Niemeyer, S., Santini, F., & van den Berg, M. (2019). Use and relevance of European Union crop monitoring and yield forecasts. *Agricultural Systems*, *168*, 224–230. https://doi.org/10.1016/j.agsy.2018.05.001

van der Velde, M., Tubiello, F. N., Vrieling, A., & Bouraoui, F. (2012). Impacts of extreme weather on wheat and maize in France: evaluating regional crop simulations against observed data. *Climatic Change*, *113*(3–4), 751–765. https://doi.org/10.1007/s10584-011-0368-2

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

Veihmeyer, F. J., & Hendrickson, A. H. (1931). The Moisture Equivalent as a Measure of the Field Capacity of Soils. *Soil Science*, *32*(3). https://journals.lww.com/soilsci/fulltext/1931/09000/the_moisture_equivalent_as_a_measure_of_the_field.3.aspx

Wallor, E., Kersebaum, K.-C., Ventrella, D., Bindi, M., Cammarano, D., Coucheney, E., Gaiser, T., Garofalo, P., Giglio, L., Giola, P., Hoffmann, M. P., Iocola, I., Lana, M., Lewan, E., Maharjan, G. R., Moriondo, M., Mula, L., Nendel, C., Pohankova, E., … Trombi, G. (2018). The response of process-based agro-ecosystem models to within-field variability in site conditions. *Field Crops Research*, *228*, 1–19. https://doi.org/10.1016/j.fcr.2018.08.021

Wang, H., Jia, G., Fu, C., Feng, J., Zhao, T., & Ma, Z. (2010). Deriving maximal light use efficiency from coordinated flux measurements and satellite data for regional gross primary production modeling. *Remote Sensing of Environment*, *114*(10), 2248–2258. https://doi.org/10.1016/j.rse.2010.05.001

Wang, J., Wang, Y., & Qi, Z. (2024). Remote Sensing Data Assimilation in Crop Growth Modeling from an Agricultural Perspective: New Insights on Challenges and Prospects. *Agronomy*, *14*(9), 1920. https://doi.org/10.3390/agronomy14091920

Weil, R. R., & Brady, N. C. (2016). *The Nature and Properties of Soils* (15th ed.). Columbus : Pearson.

Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, *236*(December 2018), 111402. https://doi.org/10.1016/j.rse.2019.111402

Wijmer, T., Al Bitar, A., Arnaud, L., & Ceschia, E. (2019). *ACEO Version 1.5.0*. CESBIO (Université de Toulouse, CNES, CNRS, UPS, IRD, INRAE).

Wijmer, T., Al Bitar, A., Arnaud, L., Fieuzal, R., & Ceschia, E. (2024). AgriCarbon-EO v1.0.1: large-scale and high-resolution simulation of carbon fluxes by assimilation of Sentinel-2 and Landsat-8 reflectances using a Bayesian approach. *Geoscientific Model Development*, *17*(3), 997–1021. https://doi.org/10.5194/gmd-17-997-2024

Wu, B., Meng, J., Li, Q., Yan, N., Du, X., & Zhang, M. (2014). Remote sensing-based global crop monitoring: experiences with China's CropWatch system. *International Journal of Digital Earth*, *7*(2), 113–137. https://doi.org/10.1080/17538947.2013.821185

Yiğit, A., & Chmielewski, F.-M. (2024). A Deeper Insight into the Yield Formation of Winter and Spring Barley in Relation to Weather and Climate Variability. *Agronomy*, *14*(7), 1503. https://doi.org/10.3390/agronomy14071503

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., Reed, B. C., & Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, *84*(3), 471–475. https://doi.org/10.1016/S0034-4257(02)00135-9

Appendix A

Table of all fixed and calibrated parameters for barley to be used in SAFYE-CO2.

Table A1

Name	Notation	Unit	Value	Method	Reference
Polynomial degree of thermal stress function	**Tpsn**		2	Literature	Duchemin (2008)
DAM on DOY of emergence	**DAM_init**	g/m2	3.3	Literature	SAFYE-CO2 documentation
DGM on DOY of emergence	**Root_init**	m	0.301	Literature	SAFYE-CO2 documentation
Root mean growth rate	**RzV**	m/°C	0.0072	Literature	SAFYE-CO2 documentation
Maximum root depth	**max_root**	m	1.55	Literature	SAFYE-CO2 documentation
Maximum vegetation fraction cover	**CovE**		0.31	Literature	SAFYE-CO2 documentation
Fraction cover exponential coefficient	**CovX**		0.91	Literature	SAFYE-CO2 documentation
Light interception coefficient	**Kex**		0.76	Literature	SAFYE-CO2 documentation
Light Use Efficiency Parameter B	**LUEb**		1.34	Literature	Pique et al. (2020a), SAFYE-CO2 documentation
Fraction of photosynthetically active radiation in Rg	**RG_par**		0.48	Literature	Pique et al (2020a), SAFYE-CO2 documentation, Varlet-Grancher (1982)
Carbon content per vegetation	**Ccont**	gC/gVeg	0.46	Literature	Pique et al. (2020a), SAFYE-CO2 documentation
Maintenance respiration parameter	**Q10**		2	Literature	Pique et al. (2020a), SAFYE-CO2 documentation, Amthor (2000)
Reference maintenance respiration at 10°C	**R10**	gC/gDAM	0.0025	Literature	Pique et al. (2020a), SAFYE-CO2 documentation, Béziat (2009)
Growth respiration conversion efficiency	**Yg**		0.74	Literature	Pique et al. (2020a), Amthor (1989)

Minimum temperature for growth	**Tmin**	°C	0	Literature	SAFYE-CO2 documentation
Maximum temperature for growth	**Tmax**	°C	37	Literature	SAFYE-CO2 documentation
Optimum temperature for growth	**Topt**	°C	20	Literature	SAFYE-CO2 documentation
Sum of temperature at which grain filling starts	**strt_grain**	°C	740	Literature	SAFYE-CO2 documentation
Corrective factor over GPP during senescence	**Cs**		1.2	Literature	Pique et al. (2020a)
Vegetation water stress soil moisture threshold	**dft**	m3/m3	0.55	Literature	SAFYE-CO2 documentation
Root water stress modifier	**SACC**		1	Literature	SAFYE-CO2 documentation
Vegetation evaporation stress parameter	**Dfe**		0.94	Literature	SAFYE-CO2 documentation
Transpiration reduction coefficient	**ktrp**		0.36	Literature	SAFYE-CO2 documentation
Maximum transpiration coefficient	**Kcb**		0.98	Literature	SAFYE-CO2 documentation
Amount off water applied per irrigation event	**irri_dose**	m/day	0	Literature	SAFYE-CO2 documentation
Minimum delay between 2 irrigation events	**irri_return_min**	Days	0	Literature	SAFYE-CO2 documentation
Soil moisture threshold at which irrigation is applied	**irri_tresh**	m3/m3	0	Literature	SAFYE-CO2 documentation
Amount of harvested grain biomass	**G_harv**		1	Literature	SAFYE-CO2 documentation
Root biomass partitioning parameter a	**PRT_Ra**		0.63	Literature	SAFYE-CO2 documentation
Root biomass partitioning parameter b	**PRT_Rb**		0.11	Literature	SAFYE-CO2 documentation
Root biomass partitioning parameter c	**PRT_Rc**		1.48	Literature	SAFYE-CO2 documentation

Specific Leaf Area modifier given thermal age	**SLAa**	m2/g°C	0	Literature	SAFYE-CO2 documentation
Emergence day of the year	**emerg**	DOY	30 - 110	LAI curve extraction	
Harvest day of the year	**harv**	DOY	150 - 250	LAI curve extraction	
LAI on DOY of emergence	**GLA_init**	m2/m2	-	LAI curve extraction	
Sum of temperature at DOY of LAI max	**SENa**	°C	-	LAI curve extraction	
Rate of senescence	**SENb**	°C/day	50,000 - 350,000	Calibration	SAFYE-CO2 documentation, Sensitivity Analysis
Amount of DAM that is not alocated to leaves at emergence	**PRT_La**		0.025-0.625	Calibration	SAFYE-CO2 documentation, Sensitivity Analysis
Sum of temperature at end of leaf allocation	**PRT_Le**	°C	100 - SENa	Calibration	SAFYE-CO2 documentation, Sensitivity Analysis
Ratio of leaf area to leaf dry mass (Specific Leaf Area)	**SLAb**	m2/g	0.006 - 0.014	Calibration	Pique et al. (2020a), SAFYE-CO2 documentation, Sensitivity Analysis
Light Use Efficiency Parameter A	**LUEa**	gC/MJ	0.95 - 1.15	Calibration	SAFYE-CO2 documentation, Sensitivity Analysis
Amount of DAM converted to grain	**PRT_G**		-	Calibration	

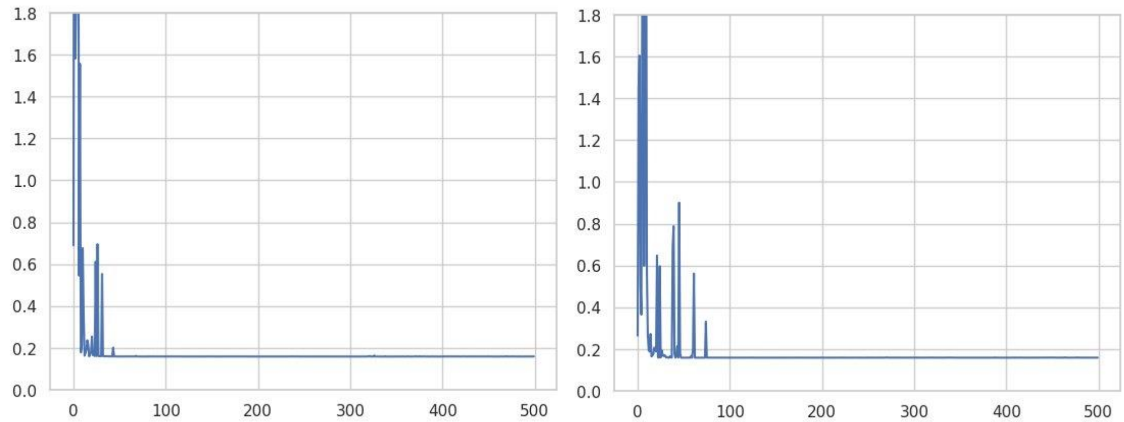
Note. SLAa was set to 0 as it was not used in previous SAFY(E)-CO2 studies either and hence no proper reference values were available. Irri_dose was set to 0 as recommended by the documentation to ensure no irrigation process is used in the modeling process.

Appendix B

Identification of required runs using Bayesian Optimization, to find the local/global minimum.

Figure B1

RMSE between observed and modeled LAI for 500 runs using Bayesian Optimization for parcel 103610 with different initial conditions.

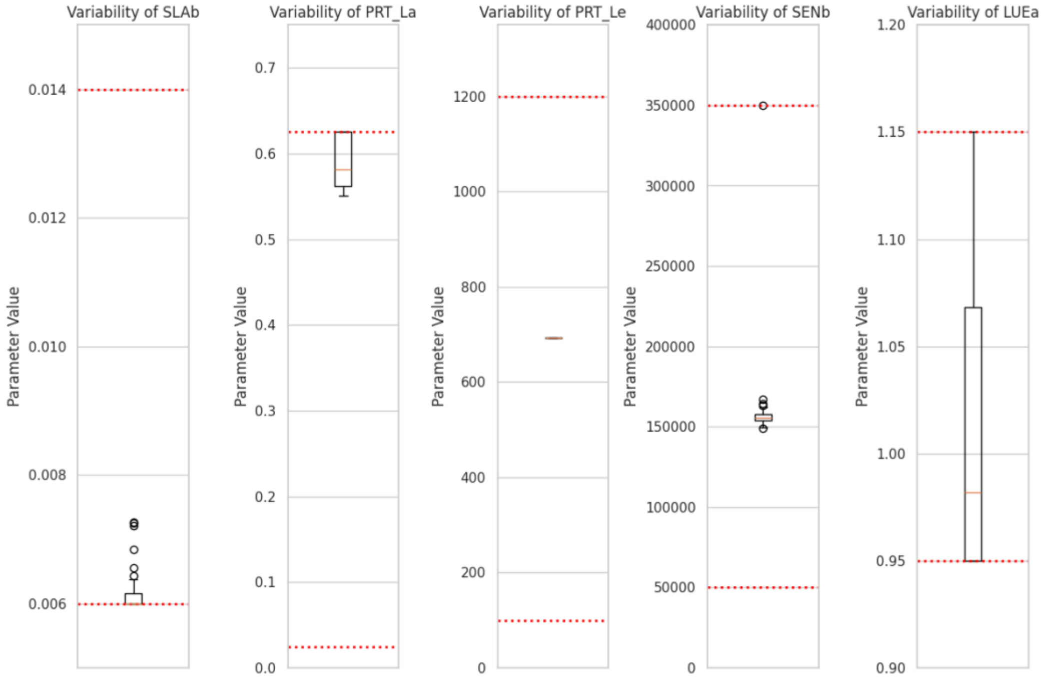


Appendix C

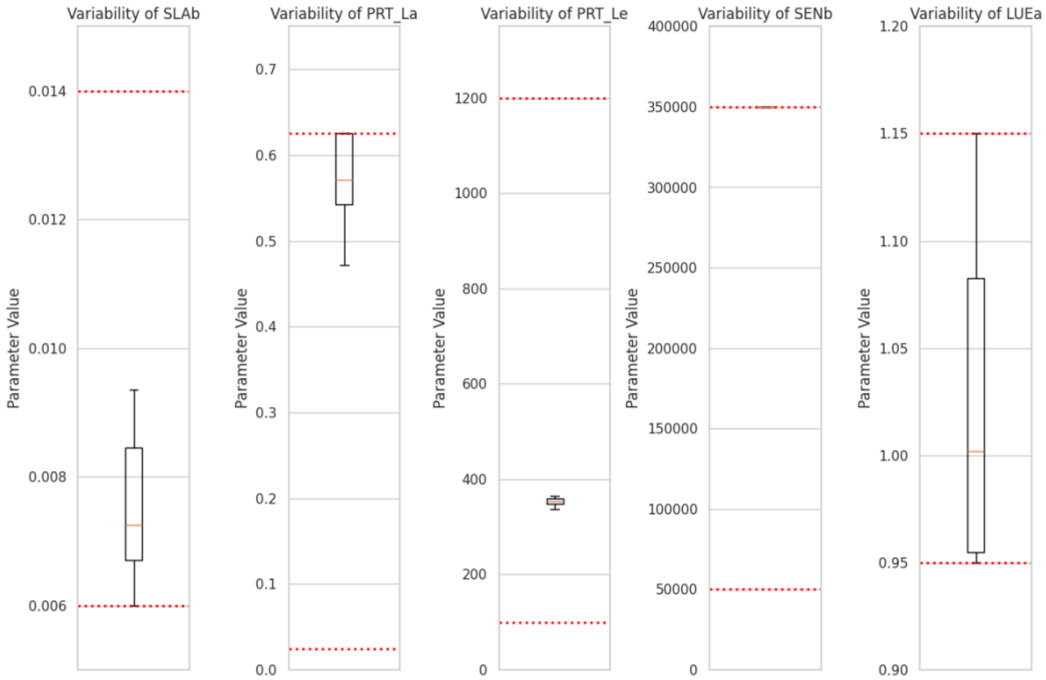
Boxplot version of the strip plots in Figure 22.

Figure C1

Calibrated parameter values for each of the 50 Bayesian Optimization calibrations for A) parcel 18595 with LAI max = 6.5 and B) 103610 with LAI max = 2.9. Colored based on the achieved RRMSE between modeled and observed LAI.



A)



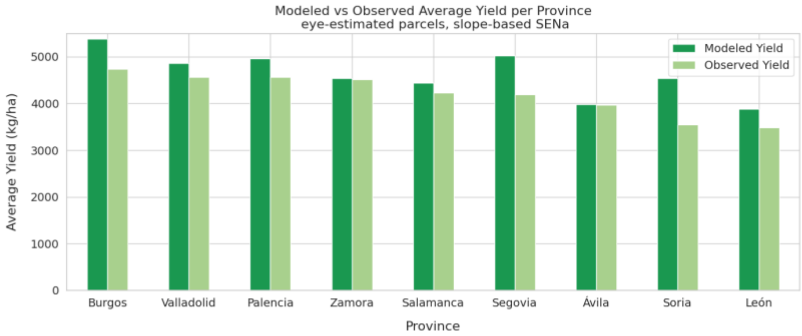
B)

Appendix D

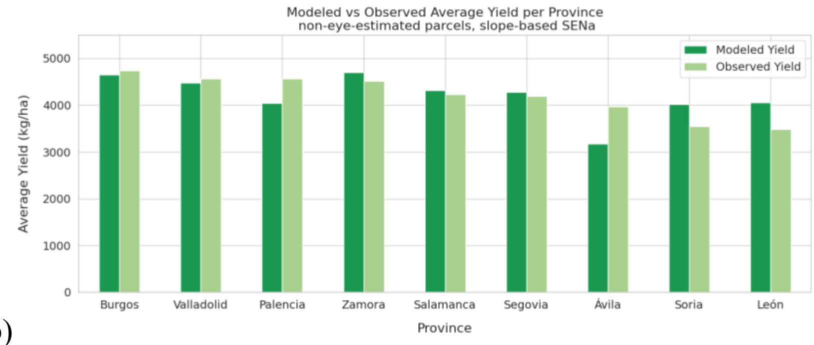
Results of provincial level yield modeling using slope-based SENa.

Figure D1

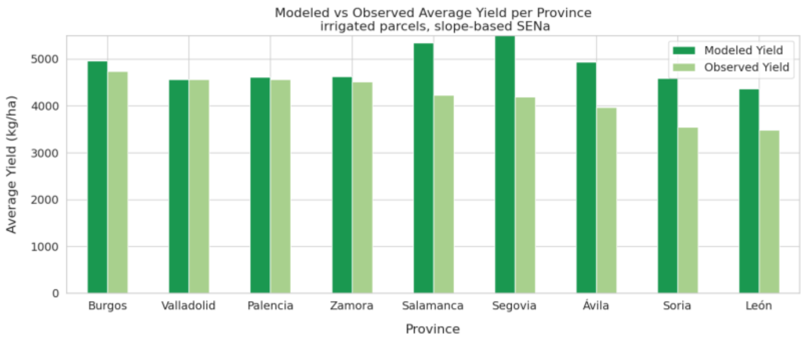
Bargraphs comparing the observed and modeled yield per province within Castilla y León. Divided into different parcel groups.



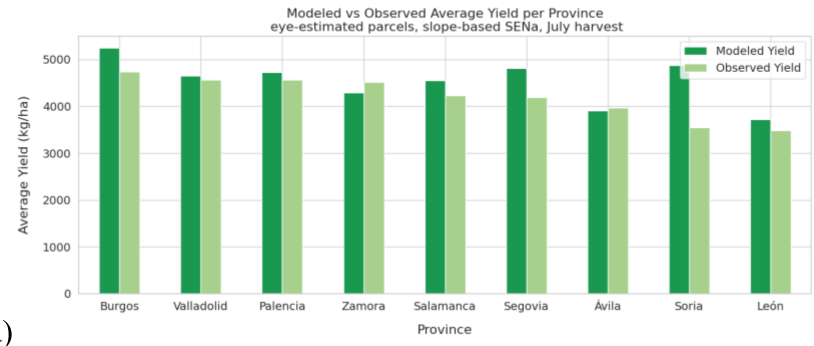
a)



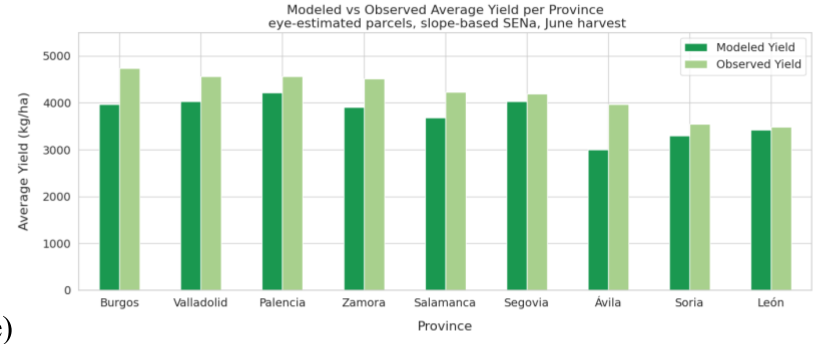
b)



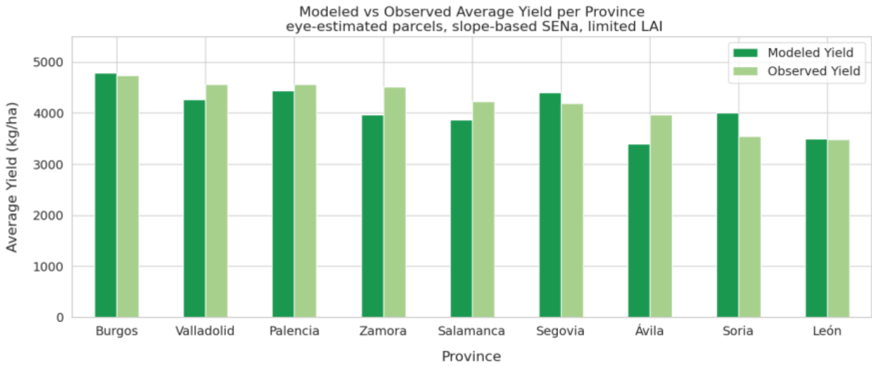
c)



d)



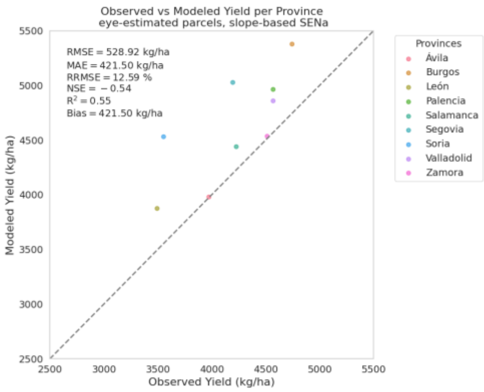
e)



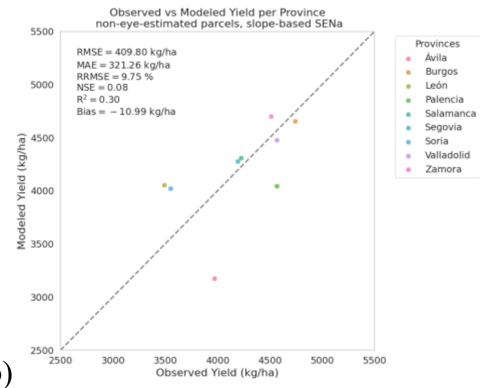
f)

Figure D2

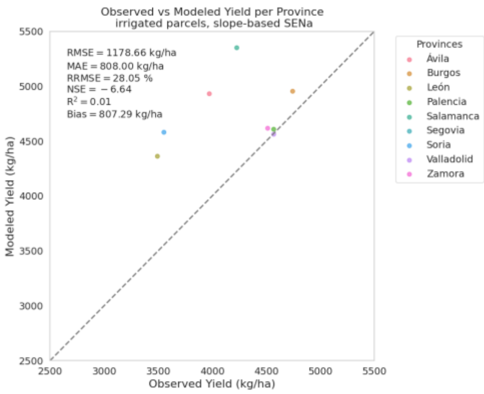
Scatterplots comparing modeled and observed yield per province in Castilla y León. Divided into different parcel groups.



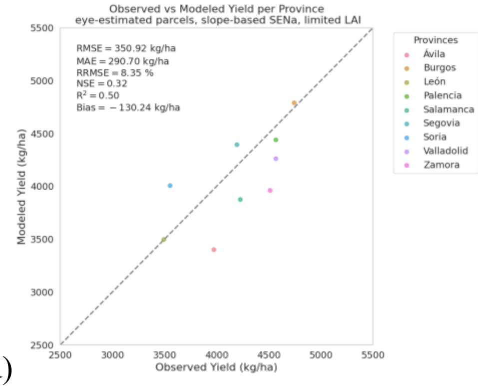
a)



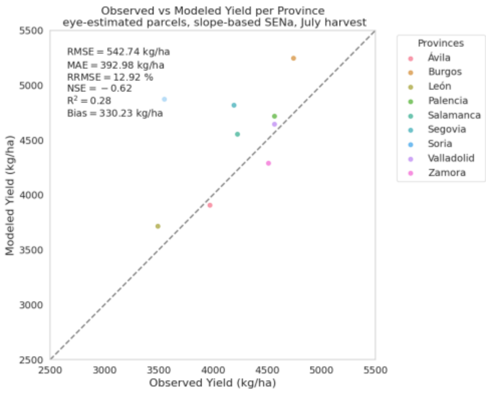
b)

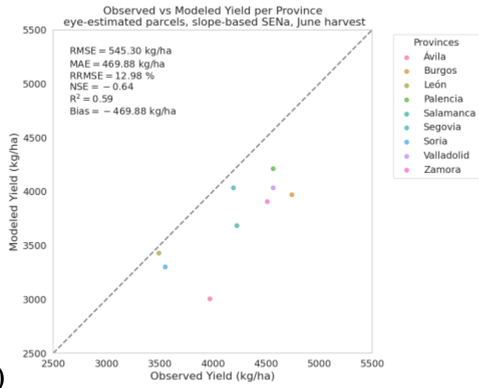


c)



d)


Observed vs Modeled Yield per Province
eye-estimated parcels, slope-based SENa, July harvest
RMSE = 542.74 kg/ha
MAE = 392.98 kg/ha
RRMSE = 12.92 %
NSE = − 0.62
R^2 = 0.28
Bias = 330.23 kg/ha
Provinces
Ávila
Burgos
León
Palencia
Salamanca
Segovia
Soria
Valladolid
Zamora
Modeled Yield (kg/ha)
Observed Yield (kg/ha)
e)

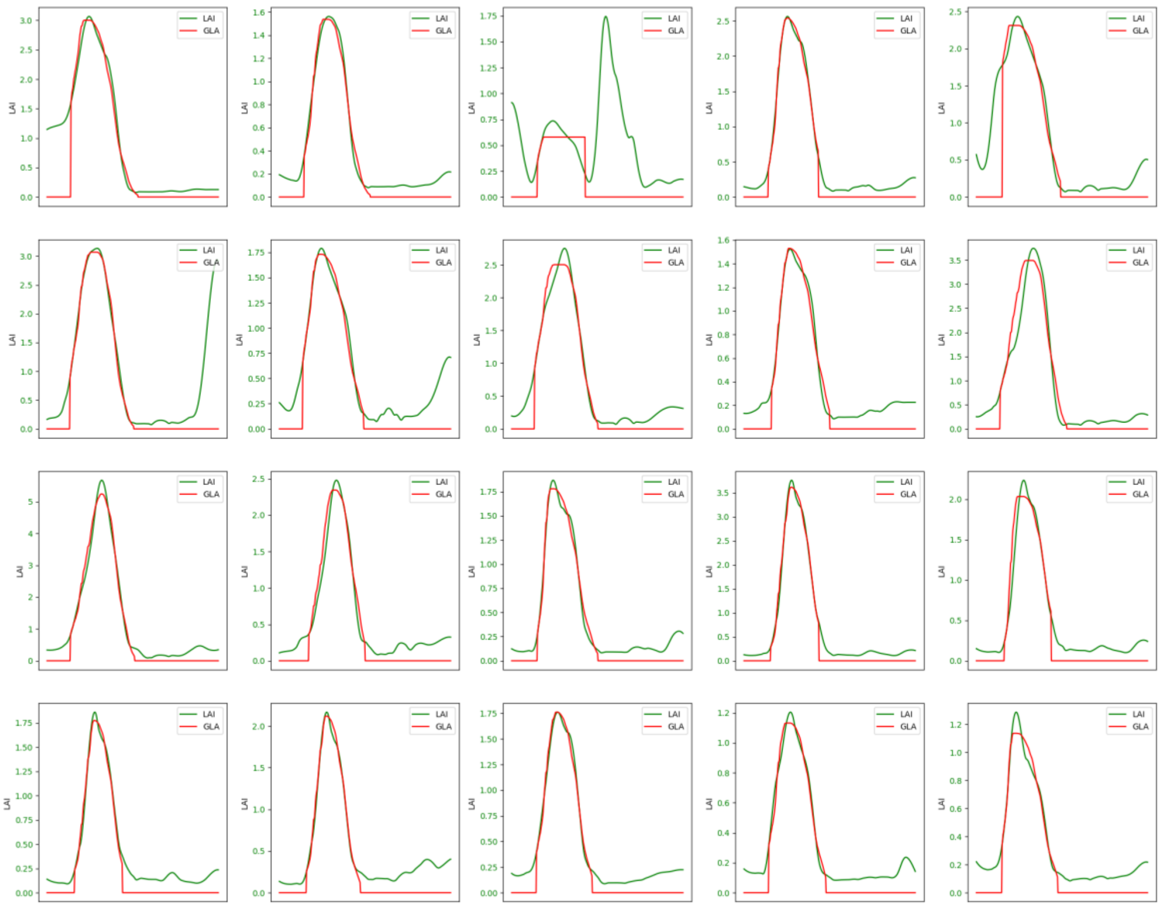

Observed vs Modeled Yield per Province
eye-estimated parcels, slope-based SENa, June harvest
RMSE = 545.30 kg/ha
MAE = 469.88 kg/ha
RRMSE = 12.98 %
NSE = − 0.64
R^2 = 0.59
Bias = − 469.88 kg/ha
Provinces
Ávila
Burgos
León
Palencia
Salamanca
Segovia
Soria
Valladolid
Zamora
Modeled Yield (kg/ha)
Observed Yield (kg/ha)
f)

Appendix E

Observed and modeled LAI curves reveal wrong interpretation of emergence and harvest for one parcel in the Àvila province for non-eye-estimated parcels.

Figure E1

Modeled and observed LAI for Àvila parcels using non-eye-estimated parcels and SENa 1 method.

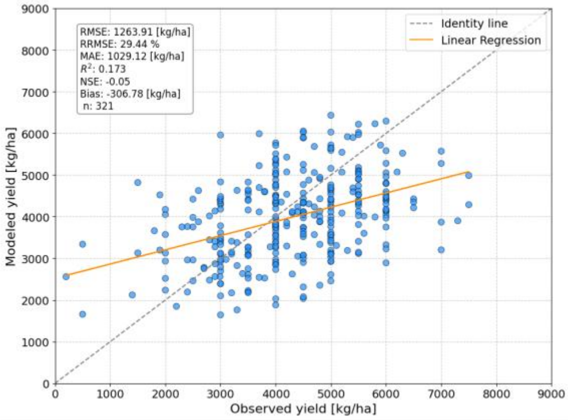


Appendix F

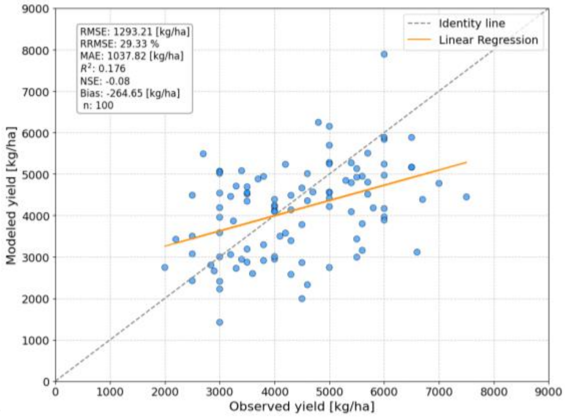
Comparison of modeled and observed yield for three parcel groups using SENa 1 method. Only displaying validation parcels, to directly compare to the Sen4Stat results.

Figure F1

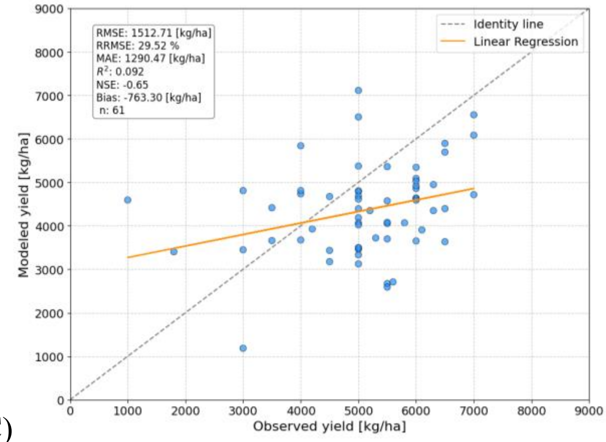
Comparison of observed to modeled yield for A) eye-estimated, B) non-eye-estimated, and C) irrigated parcels.



A)



B)



C)

Abstract

Crop yield estimates are valuable information for a wide range of stakeholders, from farmers to national governments and traders, to ensure smooth functioning of the food system and to combat food insecurity. Currently, most yield estimates are produced at national or regional scales, with governments primarily relying on cost- and time-intensive survey sampling methods, to inform agricultural policies. However, technological advances in remote sensing and modeling techniques offer opportunities to improve the accuracy and level of insight on yield estimates. Therefore, this study explores the potential of generating parcel-level crop yield estimates for barley production in Castilla y León, Spain, by applying the latest version of the process-based SAFY model series, in combination with the assimilation of the remote-sensing-derived biophysical variable Leaf Area Index (LAI). Additionally, this study presents one of the first applications of the AgriCarbon-EO (ACEO) processing chain, a new hybrid tool designed to efficiently produce yield estimation at the pixel level across large regions. The SAFYE-CO2 model was applied at the parcel level, achieving an RMSE of 1192 – 1318 kg/ha and an R^2 of 0.16 – 0.21. These results are comparable to the performance of the statistical Sen4Stat model, which yielded an RMSE of 962 – 1182 kg/ha and R^2 of 0.29 – 0.37. However, the application of the ACEO processing chain revealed considerable limitations in its ability to distinguish between parcels with low and high LAI, resulting in a lower predictive performance (RMSE: 1212 kg/ha, R^2: 0.01). The study demonstrates that while SAFYE-CO2 is capable of producing parcel-level yield estimates, improvements are needed to enhance its robustness and scalability, particularly when used in conjunction with ACEO. Key areas for improvement include the need to address equifinality issues in the calibration of SAFYE-CO2 and enhancing the parameterization and integration of remotely sensed data for ACEO. These insights contribute to ongoing efforts to develop operational, scalable tools that support parcel-level crop monitoring.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Faculté des bioingénieurs
Croix du Sud, 2 bte L7.05.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/agro**