Revisiting frameworks for crop yield prediction based on remote sensing: an agricultural perspective

by

Martin Menendez Coccoz

B.S., Universidad de Buenos Aires, 2023

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Agronomy
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2025

Approved by:

Major Professor
Ignacio A. Ciampitti

Copyright

Abstract

Crop yield predictions are essential to anticipate production volumes, reduce price volatility, and optimize logistics, among other purposes. Currently, one of the most widespread methods for building these predictions is based on the use of remote sensing, which provides spectral crop data in a rapid, precise, and non-destructive manner. Since early 1970s, multiple frameworks have emerged to integrate spectral data into crop yield prediction models. In this thesis, we revisited these frameworks and discussed their strengths and limitations. We have structured this thesis as follows. The first chapter focused on presenting the topic and main objectives of the thesis. The second chapter synthesized existing frameworks for crop yield prediction based on remotely-sensed spectral data. We described different groups and subgroups and discussed three key aspects for crop yield prediction frameworks: 1) data requirements for calibration, 2) trade-off between interpretability and predictive power, and 3) spatio-temporal transferability. The groups ranged from models directly based on remotely-sensed spectral data—whether simple or machine learning-based—to those that rely on derived biophysical variables, such as leaf area index, which can be later assimilated into crop simulation models. The third chapter applied this knowledge to a specific case, evaluating the spatio-temporal transferability of wheat (*Triticum aestivum L.*) yield prediction models at the field-scale, only requiring sowing date and location as user inputs. We used a dataset of 5,265 wheat fields in Argentina (2018–2023) including yield, sowing date, location, and model-estimated phenology. By integrating satellite-derived absorbed radiation with weather variables, we developed Elastic-Net models to predict yield from 50 days before anthesis to 20 days after. The lowest predictive error in non-transfer models occurred at anthesis, approximately 50 days before harvest (0.61 t ha^{-1}; 16%). When transferred, models maintained accuracy in unknown years, but lost accuracy in unknown regions. Additionally, we simulated operational conditions by forecasting yield in future years using models trained only on previous years, obtaining a median error of 0.67 t ha^{-1} (17%). We discussed model transfer errors in the context of absorbed radiation-to-yield conversion efficiency, which appeared to vary more across regions than years. Finally, in the fourth chapter, we provided the final remarks and offered insights to improve yield prediction models for real-world applications.

Keywords: forecast; estimation; optical data; vegetation index; NDVI

Table of Contents

List of Figures

List of Tables

Dedication

Para Feli y Juana.

1. General introduction

Predicting crop yields has been a necessity throughout history. As early as 3,000 B.C., ancient Egyptians used instruments known as "nilometers" to measure the annual flooding of the Nile, which allowed them to predict soil fertility and thus the yield of crops such as wheat and barley. These estimates were crucial for administrative purposes, as they facilitated the prediction of famines, the mitigation of flood-related disasters, and the determination of tax levels (Ardagh, 1889). Although methods have evolved, the need for accurate yield predictions remains essential (Basso & Liu, 2019).

These predictions serve as key inputs for decision-making at different levels of the agri-food supply chain (Benami et al., 2021; Fritz et al., 2019; Schauberger et al., 2020). At the regional scale, early warning systems help governments and NGOs identify vulnerable areas and optimize resource allocation. At the field scale, they enable farmers to assess the feasibility of agronomic interventions and to make logistical and commercial decisions, such as harvest planning or forward grain sales. At an even finer scale, within the field itself, yield predictions help to identify spatial variability, define management zones and increase the efficiency of input use. In addition, predictions from past growing seasons allow for the identification of yield-determining factors and the quantification of the gap between actual and potential yields (Lobell et al., 2013).

There are several methods for predicting crop yields. At the regional scale, models primarily rely on correlations between yield and weather or satellite data but often incorporate subjective information, such as expert opinion (Fritz et al., 2019). At the field scale, crop simulation models predict yield based on detailed data on weather, soil, genotype, and agronomic management (Soler et al., 2007). However, the availability of such data is often limited, which may restrict their applicability. In this context, remote sensing has gained prominence as it enables the rapid, precise, and non-destructive assessment of crop conditions, making it the main tool for yield prediction today (Basso & Liu, 2019).

Data from remote sensors can be incorporated into yield prediction models in a number of ways. These sensors rely on the interaction between incident radiation and actively growing vegetation, which absorbs and reflects radiation differently depending on the portion of the electromagnetic spectrum (Sellers, 1987). For example, vegetation absorbs most of the radiation in the visible spectrum (400-700 nm) and primarily reflects in the near infrared (750-850 nm). These optical properties allow remote sensors to be used to quantify crop condition based on vegetation indices (VIs, Zeng et al., 2022), which can be used directly in empirical models to predict yield (e.g., Bolton & Friedl, 2013) or to estimate biophysical variables that are subsequently integrated into crop simulation models (e.g., leaf area index, Huang et al., 2015). In addition, advances in programming, statistics and computing have facilitated the adoption of machine learning algorithms, which have increased predictive power (e.g., Cai et al., 2019). As a result, a wide range of remotely-sensed yield prediction frameworks have been developed in recent years, ranging from simple correlations with VIs to the assimilation of biophysical variables into simulation models, as well as complex machine learning algorithms.

Due to this diversity, the general objective of this thesis was to revisit existing frameworks for crop yield prediction using remote sensing. To achieve this, we have structured this thesis into the following chapters, each aligning with specific objectives. In the second chapter, our objective was to provide a synthesis on existing frameworks for crop yield prediction based on remotely-sensed spectral data. We described and exemplified the main groups of frameworks and discussed three fundamental aspects common to all: the data required for model calibration, the trade-off between interpretability and predictive power, and their spatio-temporal transferability. In the third chapter, we applied these concepts to a specific case study. Here, our objective was to assess the spatio-temporal transferability of field-scale wheat yield prediction models that require only the sowing date and location as user inputs. To this end, we used a dataset of 5,265 wheat fields in Argentina (2018-2023) to develop Elastic-Net models driven primarily by satellite-derived absorbed radiation, supplemented by weather inputs. Then, we transferred these models to unknown regions and years and assessed their predictive performance. Finally, in the fourth chapter, we summarized the main findings of the thesis and discussed future research challenges related to remotely-sensed crop yield prediction models.

2. Revisiting frameworks for crop yield prediction with remotely-sensed spectral data

Abstract

Crop yield predictions are a key input for decision-making across stakeholders in the agri-food supply chain. Currently, these predictions mostly rely on remotely-sensed spectral data. The integration of these data into predictive models began in the 1970s, leading to the development of multiple frameworks, each with its own strengths and limitations. In this study, we provide a comparative synthesis of these frameworks. We identified three main groups: (1) models based directly on spectral data, mainly vegetation indices, subdivided into simple statistical models and machine learning models; (2) models based on derived biophysical variables, distinguishing between empirical models and based on resource capture and use efficiencies; and (3) models integrating spectral data with crop simulation models, including data assimilation, metamodel calibration, and hybrid approaches. These frameworks were analyzed in terms of data requirements, interpretability, predictive power, and spatio-temporal transferability. Rather than advocating for a single best framework, we emphasize that the optimal choice depends on research objectives, available data, and expertise. This structured synthesis provides insights into the strengths and limitations of different approaches, helping guide future improvements in crop yield prediction models.

Introduction

Accurate crop yield predictions are crucial for achieving food security, and multiple methods have been explored for this purpose. At regional scale, governments combine field sampling and farmer survey data to estimate yield before harvest (Nandram et al., 2014). At this scale, yield has also been correlated with weather variables (Nichols, 1985; Thompson, 1969). At field scale, crop simulation models predict yield based on detailed soil, weather, genotype, and management data (Chipanshi et al., 1997; Soler et al., 2007) and can be adapted to regional scales by assuming representative conditions (Manivasagam & Rozenstein, 2020). Today, remote sensing is the most widely used tool at both field- and regional-scales, and its integration with other data sources, such as weather or simulations models, enables the development of different frameworks for yield prediction (Box 2.1; Basso and Liu, 2019; Fritz et al., 2019; Huang et al., 2015; Lobell et al., 2015; Paudel et al., 2022).

Optical remote sensors enable the estimation of crop status and yield prediction. These sensors rely on the unique optical properties of actively growing vegetation, which interacts with radiation across different regions of the electromagnetic spectrum. For instance, vegetation absorbs a large proportion of radiation in the visible spectrum (400–700 nm), while it reflects a high proportion in the near-infrared (750–850 nm). By measuring reflectance, defined as the proportion of incident radiation that is reflected, these sensors facilitate the estimation of biophysical crop variables such as leaf area index (LAI) or the fraction of photosynthetically active radiation absorbed by the green tissues (fAPAR, Gitelson et al., 2002, 2004; Sellers, 1987; Weiss et al., 2020). These estimations can be made using vegetation indices (VIs), which are mathematical transformations of these reflectance values (Zeng et al., 2022). The most widely used VI is the Normalized Difference Vegetation Index (NDVI), which utilizes red and near-infrared reflectance (Rouse et al., 1974).

Thus, mounted on handheld devices, ground vehicles, drones, or satellites, optical remote sensors enable the precise, rapid, and non-destructive quantification of crop status.

However, how can yield be predicted from remotely-sensed spectral data? Different frameworks have emerged since the first initial efforts over four decades ago (Hammond, 1975; Heilman et al., 1977; MacDonald & Hall, 1980). From one side, there is a group of crop yield prediction frameworks based on direct correlations with VIs; while on the other hand, there is group that predicts yield via translating spectral data into biophysical variables, such as LAI or fAPAR, which can be then assimilated into crop simulation models. A diverse degree of combinations exists in between these two approaches. Thus, from simple regressions between yield and NDVI to the assimilation of biophysical variables into a simulation model, different frameworks guided by agronomical knowledge have emerged (Huang et al., 2015; Magney et al., 2016; Lobell et al., 2015). In parallel, advances in statistics, programming, and handling of large geospatial datasets have enabled the development of models that do not necessarily incorporate deep agronomical knowledge, as algorithms learn directly from the data. Today, artificial intelligence algorithms can learn the temporal evolution of VIs, independently identify the key crop phenological periods and the most relevant combinations of index and other variables for yield prediction (Cao et al., 2021; Nevavuori et al., 2019; Sun et al., 2019; Zhang et al., 2019).

After more than 40 years of research, several frameworks for yield prediction are now available. Previous efforts have summarized the state of the art in the field, but focused on multiple data sources (Basso & Liu, 2019; Schauberger et al., 2020) or specific statistical methods (Chlingaryan et al., 2018; Muruganantham et al., 2022). More than a decade ago, Rembold et al. (2013) classified remote satellite-based yield prediction frameworks into three groups: (1) regression models based on VIs; (2) models based on radiation capture and use efficiencies; and

(3) models based on the assimilation of satellite data into crop simulations. However, new approaches and technologies have emerged in the past decade (Cai et al., 2019; Gaso et al., 2024; Lobell et al., 2015; Paudel et al., 2021), underscoring the need to revisit these frameworks and assess their strengths and limitations (Maestrinni et al., 2022).

Thus, the objective of this study was to synthetize existing frameworks for crop yield prediction using remotely-sensed spectral data. Therefore, we structured the study into three specific objectives related to: (1) identify and describe different framework groups (and subgroups) and how they differ in their use of spectral data. While we define groups for clarity, we acknowledge that these frameworks exist on a continuum and may overlap; (2) discuss data requirements for model calibration, the trade-off between interpretability and predictive power, and the spatio-temporal transferability of different frameworks; and (3) highlight pathways for improving crop yield prediction models. This study did not aim to identify a single best framework, with the optimal choice depending on the research objective, data availability, and level of expertise.

Box 2.1: What do we understand by yield "prediction", and how does the scale affect prediction errors?

The term "prediction" is often used interchangeably to refer to different approaches aimed at quantifying crop yield. Here, we distinguish two primary objectives: (1) forecasting, which estimates yield before harvest, and (2) diagnosis (or hindcasting), which estimates yield after harvest. Forecasting aims to anticipate yield to support tactical decision-making, such as logistical planning or crop management. It relies solely on information available up to the forecast time (e.g., data available up to flowering). Accuracy improves as the crop cycle progresses, incorporating more detailed and relevant information. In early stages, forecasts often rely on historical yield averages, eventually adjusted for the environmental context. In later stages, uncertainty decreases as more information is included, particularly when the prediction occurs during or after the crop's critical period (Bolton & Friedl, 2013; Schwalbert et al., 2020). In contrast, diagnosis uses all available information up to harvest, resulting in lower errors compared to early forecasts. Its primary value lies in analyzing historical yield variability to identify the influence of environmental and management factors, and quantify gaps relative to yield potential (Deines et al., 2023, 2024; Lobell et al., 2013).

The accuracy of yield prediction models also depends on the scale of application. At coarser scales, reduced variability and the aggregation of predictions from smaller units tend to lower errors. For example, models calibrated at regional scales typically exhibit relatively low errors, around 10% (Becker-Reshef et al., 2010; Chipanshi et al., 2015; Han et al., 2020; Ren et al., 2008). In contrast, at field or subfield scales, errors are often larger, ranging from 15–35% (Chen et al., 2020; Deines et al., 2021; Ma et al., 2024). Despite the growing need for accurate field-scale models to support on-farm decision-making, most published models have been calibrated at regional scales using official yield statistics (Schauberger et al., 2020). Consequently, efforts to develop models with high-spatial resolution and broad temporal coverage at the field or subfield scale remain limited, with only a few recent exceptions (e.g., Brinkhoff et al., 2024; Deines et al., 2021; Ma et al., 2024).

Description of the different frameworks for crop yield prediction

In this section, we present the classification of frameworks available for crop yield prediction based on remotely-sensed spectral data (Fig. 2.1). Broadly, we described three main groups: (1) models based directly on spectral data; (2) models based on derived biophysical variables; and (3) models based on the integration of spectral data and crop simulation models. Within each group, we described two or three subgroups of models.

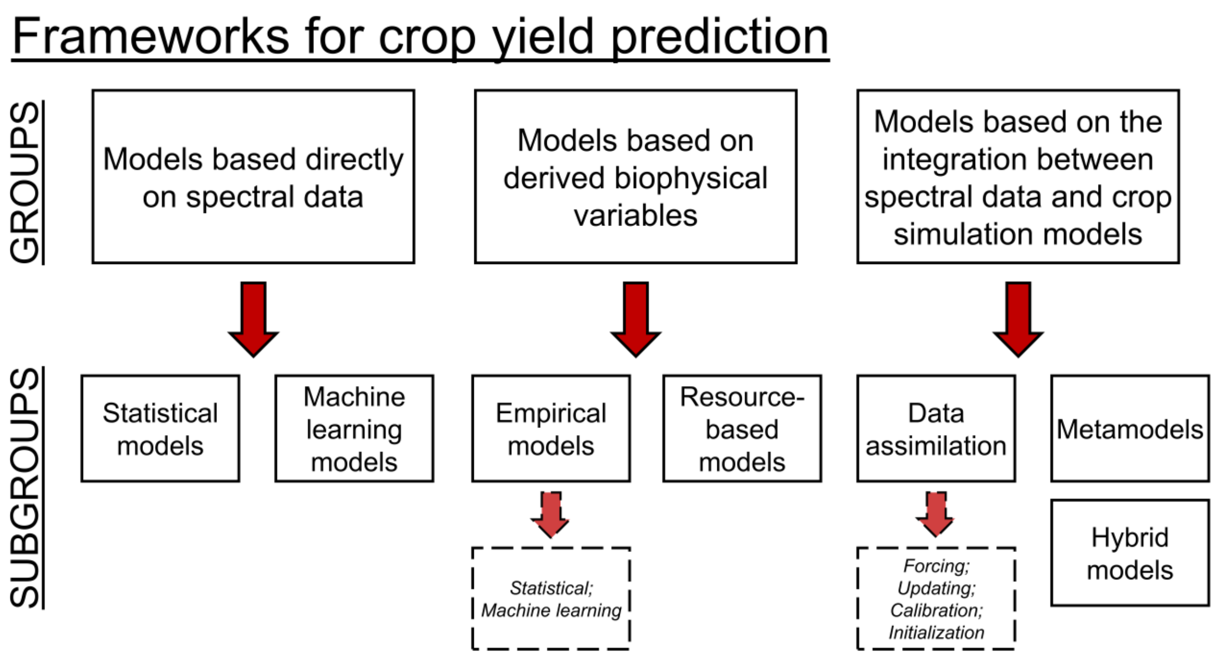

Figure 2.1: *Conceptual representation of the groups and subgroups of frameworks for crop yield prediction.*

Crop yield prediction models based directly on spectral data

In this framework group, spectral data—mainly VIs—are directly used in empirical, data-driven models to predict yield. Below, we describe two subgroups within this framework, distinguished by the complexity of the algorithm used to relate yield to spectral data (Fig. 2.2, Table 2.1).

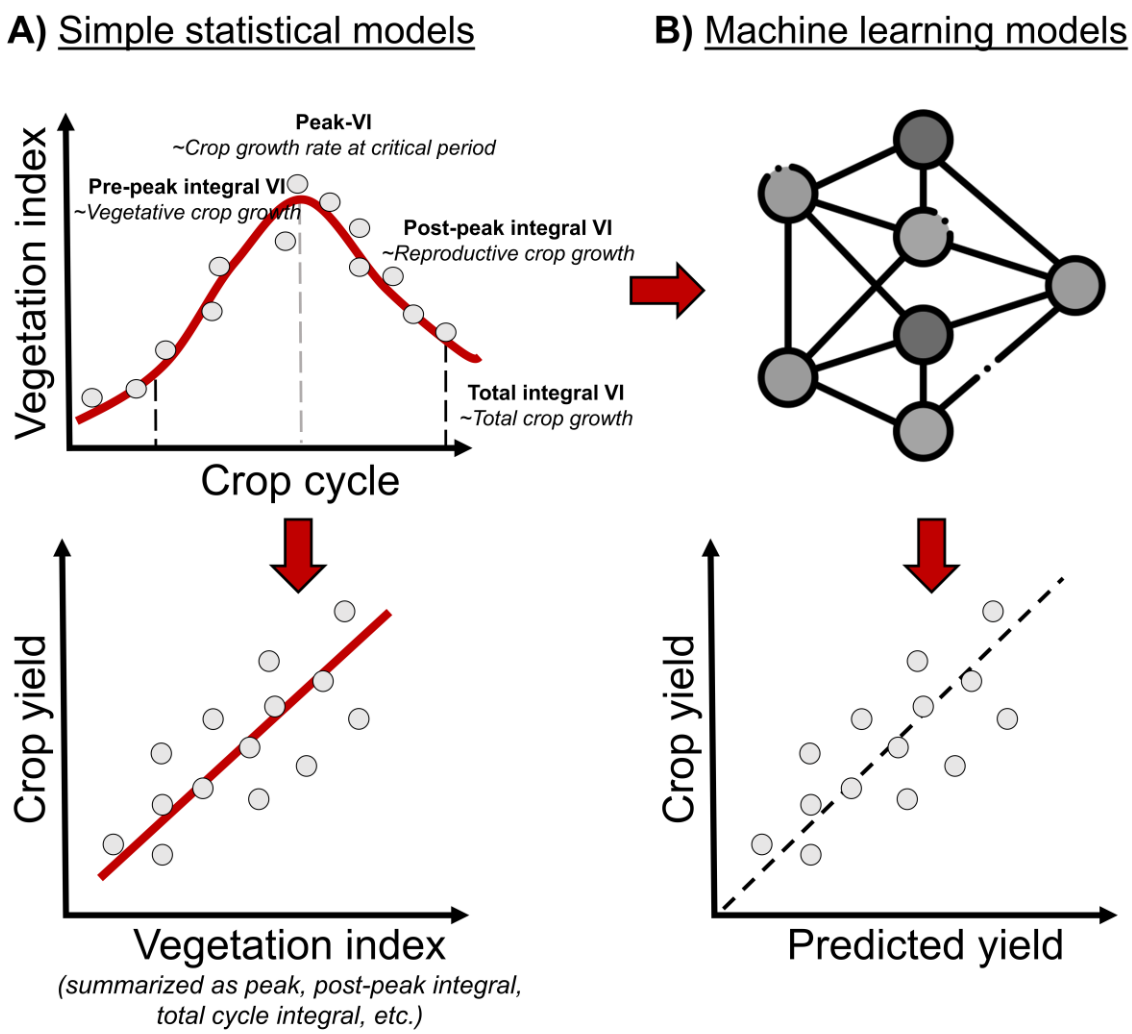

Figure 2.2: *Conceptual representation of crop yield prediction models based directly on spectral data. Subgroups: A) Simple statistical models; B) Machine learning models.*

Table 2.1: Examples of papers with models based directly on spectral data, statistical and machine learning subgroups.

Subgroup	Algorithm	Spectral data	Crop	Reference
Statistical	Linear regression	NDVI; RVI	Wheat	Tucker et al. (1981)
	Linear regression	NDVI	Millet	Groten et al. (1993)
	Linear regression	NDVI	Wheat	Ren et al. (2008)
	Linear regression	NDVI; EVI2; NDWI	Maize; Soybean	Bolton & Friedl (2013)
	Linear regression	NDVI; RVI	Wheat	Wang et al. (2014)
	Linear regression	NDVI	Wheat	Magney et al. (2016)
	Linear regression	NDVI; DVI; CIrededge; CIgreen; NDRE; EVI; EVI2; MTCI	Rice	Duan et al. (2021)
Machine learning	LASSO; RandomForest; Support Vector Machines; Neural Networks	EVI; SIF	Wheat	Cai et al. (2019)
	Neural Networks	Raw reflectance bands: Blue, Green, Red, Near Infrarred, Shortwave Infrarred	Soybean	Sun et al. (2019)
	RandomForest	EVI	Wheat; Barley; Canola	Filippi et al. (2019)
	LASSO; RandomForest; XGBoost; Neural Networks	EVI	Maize	Zhang et al. (2019)
	Linear regression; RandomForest; Neural Networks	NDVI; EVI	Soybean	Schwalbert et al. (2020)
	LASSO; RandomForest; Neural Networks	EVI	Rice	Cao et al. (2021)
	Ridge; RandomForest; Neural Networks	CIgreen	Maize; Soybean; Wheat	Ma et al. (2024)

Simple statistical models

This subgroup of models predicts yield directly from spectral data using simple statistical relationships (Fig. 2.2A, Hassan et al., 2019; Johnson, 2014; Lai et al., 2018, Mkhabela et al., 2011; Ren et al., 2008). The underlying assumption is that most environmental, management, and genetic factors that influence yield will also influence crop greenness, which can be quantified through spectral data, primarily VIs. This framework is likely the oldest, with studies dating back 40-to-50 years with yield-VIs regressions based on red and near-infrared reflectance, such as NDVI, summarized at different crop phenological stages, including peak values or integrals (Colwell et al., 1977; Groten et al., 1993; Tucker et al., 1981). More recently, additional VIs incorporating other reflectance bands, such as blue (~450 nm, EVI), green (~550 nm, GNDVI, GCVI), or red edge (~750 nm, NDRE), have been adopted. These VIs are generally less affected by soil reflectance and saturation at high LAI (Bolton & Frield, 2013; Duan et al., 2021; Peralta et al., 2016; Shanahan et al., 2001; Zhao et al., 2007). Below, we provide some examples of this subgroup of models.

At the experimental plot scale, Magney et al. (2016) used NDVI from handheld sensors and identified anthesis as the best phenological stage for wheat yield prediction ($R^2 = 0.45$), improving the model by including the duration of tillering and stem elongation phases derived from inflection points of the NDVI time series curve ($R^2 = 0.83$). A similar approach was applied by Wang et al. (2014), who used SPOT-5 and HJ-CCD satellite imagery to analyze wheat from tillering to maturity, with the most accurate prediction model integrating NDVI and RVI between stem elongation and early grain filling ($R^2 \sim 0.65$), both indices relying on red and near-infrared reflectance. Following this regression-based approach, Bolton & Friedl (2013) designed a system to forecast county-scale maize and soybean yield in United States. They calculated NDVI, EVI2, and NDWI from MODIS satellite, and found that the optimal timing for yield forecasts corresponded to the peak of each index—which coincided with both critical periods for these crops (Gao et al., 2017). Moreover, they determined that in arid regions, NDWI, an index more sensitive to vegetation water content, was the most effective, while in other regions EVI2 performed best.

Machine learning models

This subgroup of models leverages machine learning algorithms to capture complex, nonlinear relationships between yield and multiple predictors (Fig. 2.2B; Filippi et al., 2019; Kamir et al., 2020; Ma et al., 2021, 2024; Nevavuori et al., 2019; Sun et al., 2019; Zhang et al., 2019). According to bibliometric analyses, studies with these models have increased exponentially since 2018–2019 and currently dominate the field (Clark et al., 2023). These models often integrate spectral data with additional sources, typically gridded weather and soil datasets. As a result, they incorporate multiple predictor variables that are often correlated (e.g. two or more VIs derived from the same reflectance bands). A typical study in this framework compares the importance of different data sources, such as spectral, weather or soil data, for yield prediction, along with the accuracy of different machine learning algorithms (e.g., Cai et al., 2019; Cao et al., 2021; Schwalbert et al., 2020). A few examples from the literature are illustrated below.

Cai et al. (2019) forecasted wheat yield at the county scale in Australia (2000–2014), evaluating different algorithms: LASSO, Random Forest, Support Vector Machines, and Neural Networks, with the highest predictive accuracy using Support Vector Machines, enabling yield forecasts two months before harvest ($R^2 \sim 0.75$). They also found that EVI, a reflectance-based VI, outperformed SIF (Solar-Induced Fluorescence), an emission-based VI. The authors attributed this to the lower spatio-temporal resolution of SIF data compared to MODIS-derived EVI. Similarly, in Brazil, Schwalbert et al. (2020) applied deep learning to soybean yield forecasting, integrating

NDVI and EVI from MODIS with precipitation and temperature data at different crop stages. Their model achieved low error at the county scale, even early in the season (mean error: 0.42 Mg ha^{-1}), improving as the season progressed (0.24 Mg ha^{-1}).

Since these algorithms require large volumes of data, they typically operate at a regional scale, as data at the field or subfield scales tend to be more limited. To overcome this limitation, Ma et al. (2024) employed transfer learning, a deep learning technique that enables models to be trained in data-rich contexts and subsequently adapted to data-scarce scenarios without compromising accuracy. Using this approach, they trained models with only county-scale data and applied them at the subfield scale (30 m), using more than 20 predictor variables derived from satellite and weather data and achieving yield prediction errors of 20% for corn, 25% for soybeans, and 34% for wheat. This methodology facilitated the generation of annual yield maps with a spatial resolution of 30 m across the United States, offering an approach that can be extrapolated to other regions where only county-scale data are available.

This section provides only a brief overview, as machine learning-based yield prediction is a much broader field, encompassing aspects such as data pre-processing, algorithm selection, and hyperparameter tuning, which are covered in specific reviews (e.g., Chlingaryan et al., 2018; Muruganantham et al., 2022; von Klomperburg et al., 2020).

Crop yield prediction models based on derived biophysical variables

In this framework group, spectral data are first used to derive biophysical variables, such as LAI or fAPAR, which are then utilized to predict yield (Fig. 2.3). This derivation can be achieved through two main approaches: (1) in situ calibration between biophysical variables and spectral data (e.g., Nguy-Robertson et al., 2012) or (2) inversion of radiative transfer models (e.g., Darvishzadeh et al., 2008). For a more detailed discussion on deriving biophysical variables from spectral remotely-sensed data, readers are referred to review studies published by Homolová et al. (2013) and Weiss et al. (2020). Below, we describe two subgroups of models that fall within this framework (Fig. 2.3, Table 2.2).

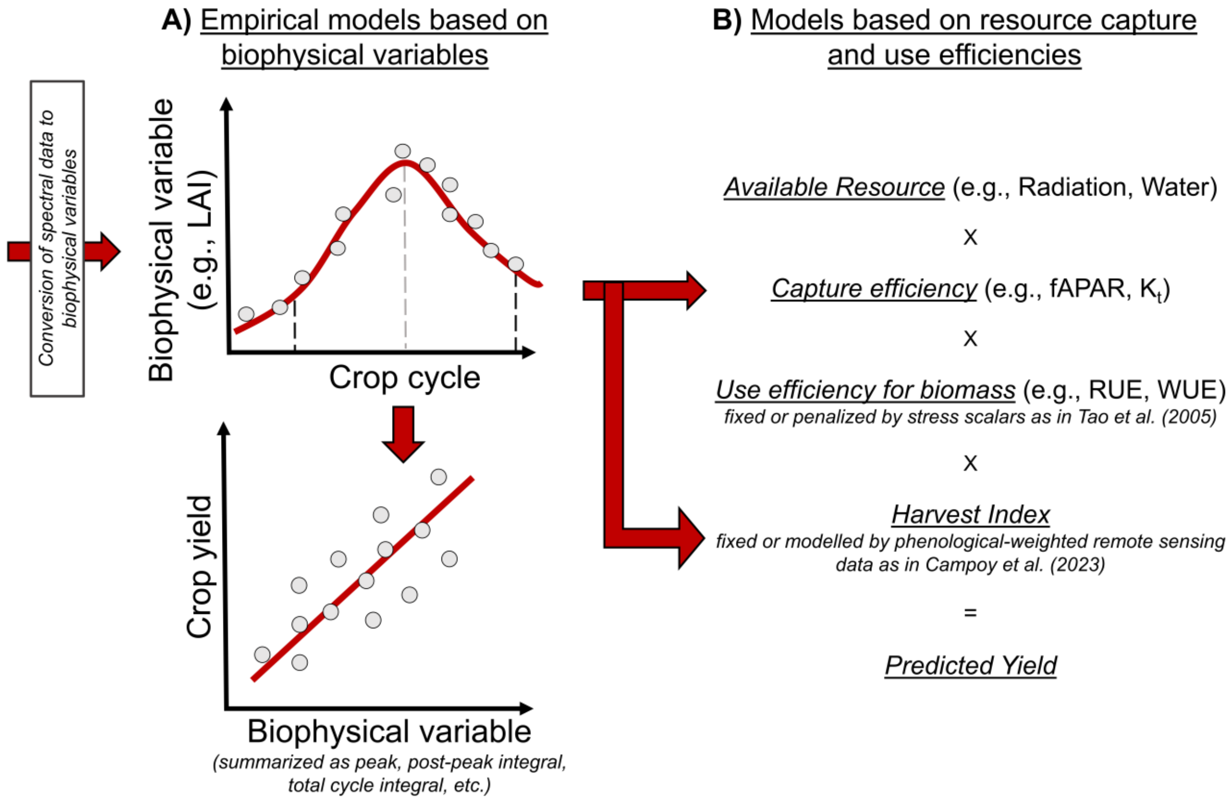

Figure 2.3: *Conceptual representation of crop yield prediction models based on biophysical variables. Subgroups: A) Empirical models; B) Models based on resource capture and use efficiencies.*

Table 2.2: Examples of papers with models based on derived biophysical variables, empirical and resource-based subgroups.

Subgroup	Algorithm	Biophysical data	Crop	Reference
Empirical	Linear regression	LAI as input to estimate evapotranspiration	Wheat	Heilman et al. (1977)
	Linear regression	fAPAR as input to estimate total absorbed radiation	Maize	Gallo et al. (1985)
	Linear regression	fAPAR	Wheat; Barley; Maize	Lopez-Lozano et al. (2015)
	Neural Networks	LAI	Wheat	Wang et al. (2022)
	Neural Networks	fAPAR; LAI	Wheat	Wang et al. (2023)
Resource-based	Radiation capture and use efficiency	fAPAR	Wheat	Lobell et al. (2003)
	Radiation capture and use efficiency	fAPAR	Wheat; Rice; Cotton; Sugarcane	Bastianseen & Ali (2003)
	Radiation capture and use efficiency	fAPAR	Wheat; Potato; Rapeseed; Soybean; Sugar Beet; Barley; Rice; Maize	Yuan et al. (2016)
	Radiation capture and use efficiency	fAPAR	Barley; Canola; Wheat	Chen et al. (2020)
	Radiation and Water capture and use efficiency	fAPAR; Kt	Wheat; Barley	Campoy et al. (2023)

Empirical models based on biophysical variables

A first subgroup of models directly regresses crop yield against biophysical variables using either simple statistical methods or machine learning algorithms (Fig. 2.3A, Gallo et al., 1985; Heilman et al., 1977; Johnson, 2016; Kolotii et al., 2015; Wang et al., 2022, 2023). For example, López-Lozano et al. (2015) predicted yield of different crops (wheat, barley, and maize) across European administrative units (1999–2012) with linear regressions based on the fAPAR product from the SPOT-VEGETATION satellite. For each administrative unit, the authors identified the temporal window of the fAPAR with the highest R^2, typically centered around one month from the peak of fAPAR. They found that regions with high interannual variability in water availability exhibited high R^2 values (~0.8), while those without water deficit showed lower R^2 values (<0.6) for yield prediction.

Models based on resource capture and use efficiencies

A second subgroup consists of models based on resource capture and use efficiencies, using biophysical variables as inputs (Fig. 2.3B, Bastiaanssen & Ali, 2003; Chen et al., 2020; Prince, 1991; Yu et al., 2024; Yuan et al., 2016; Zheng et al., 2016). The underlying logic is summarized in the following pair of equations (Monteith, 1972):

$$Yield = Biomass * Harvest\ Index \qquad (Eq. 2.1)$$

$$Biomass \ = \int_{\mathrm{em.}}^{\mathrm{mat.}} Resource * Capture\ Efficiency * Use\ Efficiency \qquad (Eq. 2.2)$$

where yield is the product of total crop biomass and the harvest index (fraction of biomass allocated to harvestable organs, Eq. 2.1). In turn, biomass is the product of the available resource between crop emergence and physiological maturity (e.g., incident photosynthetically active radiation or available water), the capture efficiency (proportion of the resource captured), and the use efficiency (biomass produced per unit of resource captured, Eq. 2.2). Spectral data allows for the derivation of capture efficiency (e.g., fAPAR for radiation, Kt for water), while the available resource can be obtained through field measurements, weather stations, or gridded databases. For example, Lobell et al. (2003) used Landsat satellites to derive wheat fAPAR and, incorporating data on incident photosynthetically active radiation, radiation use efficiency, and a fixed harvest index, accurately estimated total wheat production in Mexico at the regional scale (~3% error in the 1993–94 and 1999–2000 growing seasons).

The main challenge for this subgroup lies in accurately modeling both use efficiency and harvest index, as they vary with genotype, management, and environmental factors (Sinclair & Muchow, 1999; Unkovich et al., 2010) and cannot be directly derived from remote sensing. With this in mind, Tao et al. (2005) predicted maize yield across 122 experimental stations in China using satellite-based fAPAR and dynamically incorporating radiation use efficiency (RUE) through "scalars" that penalized potential RUE based on temperature, water, or nutrient stress (but using a constant value for harvest index). More recently, Campoy et al. (2023) predicted wheat and barley yield at the field and subfield scales using radiation and water capture and use efficiency models. With satellite NDVI, they estimated fAPAR and Kt, and accumulated absorbed radiation and transpired water over the crop cycle. Based on locally calibrated biomass production use efficiencies, they estimated total biomass. Finally, they curvilinearly related the harvest index to the proportion of the resource captured post-anthesis relative to the total cycle (e.g., the ratio between post-anthesis absorbed radiation and total absorbed radiation, Kemanian et al., 2007; Sadras & Connor, 1991). This resulted in a simple, agronomically interpretable model that predicted yield at the field and subfield scales.

Crop yield prediction models based on the integration of spectral data and crop simulation models

In this framework group, yield predictions are built by integrating spectral data with crop simulation models. Briefly, these models simulate crop growth and development through mathematical equations that quantify crop physiological processes driven by soil, weather, genotype, and agronomical management inputs. They use parameters to describe system characteristics, such as leaf expansion rate or potential radiation use efficiency. With these inputs and parameters, they simulate state variables of the system at a given time step, such as LAI or soil water content. Since the accuracy of these models depends on the quality of input data and their correct parameterization, spectral data from remote sensors provide complementary information to improve and scale yield predictions. Below, we describe three common subgroups of models that differ in how they integrate spectral data with crop simulation models (Fig. 2.4, Table 2.3).

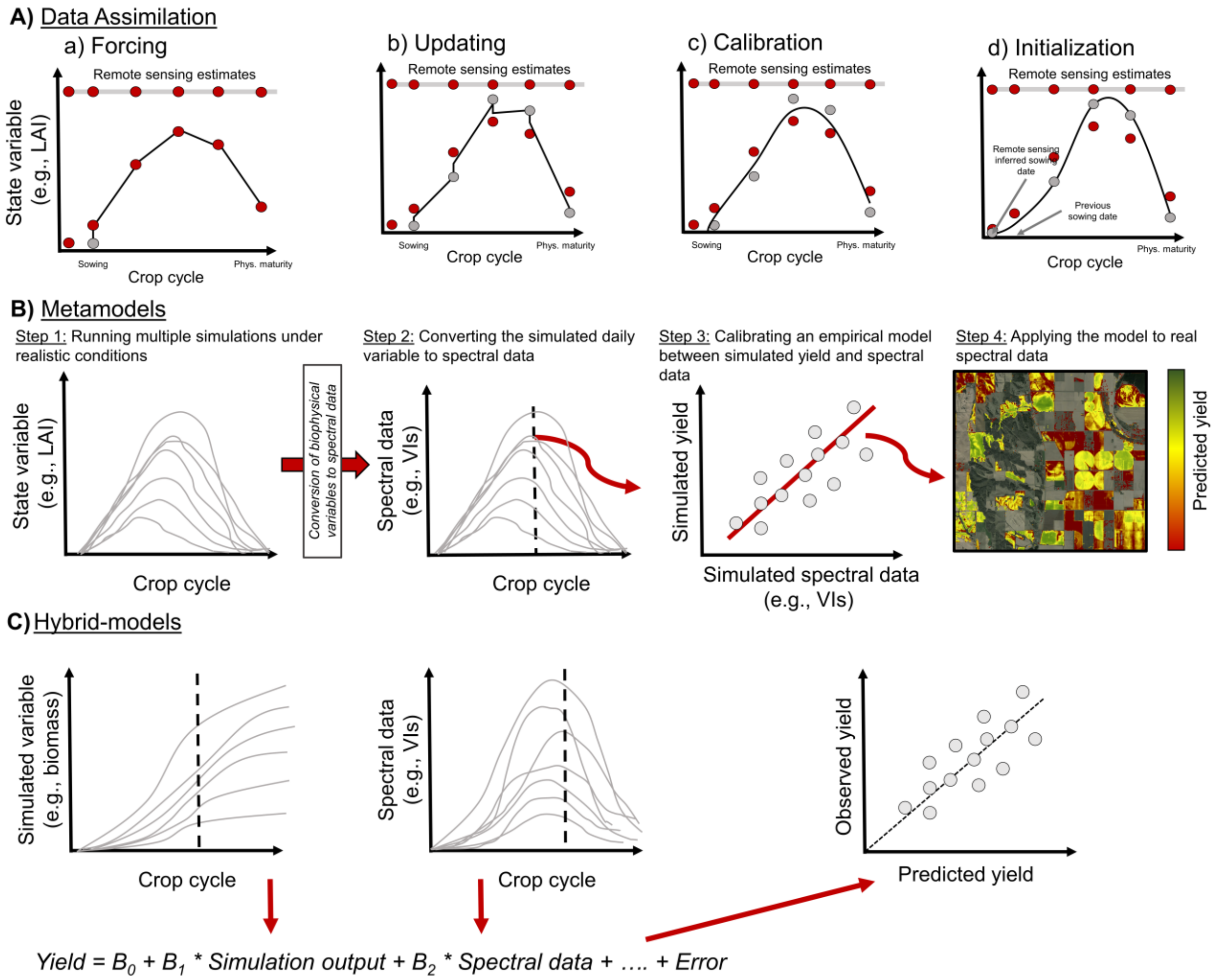

Figure 2.4: *Conceptual representation of models based on the integration of spectral data and crop simulation models. Subgroups: A) Models based on data assimilation, with four strategies: a) Forcing; b) Updating; c) Calibration; and d) Initialization; B) Metamodels based on multiple simulations; C) Hybrid models with simulated outputs and spectral data as individual predictive variables.*

Table 2.3: Example of papers with models based on the integration of spectral data and crop simulation models.

Subgroup	Strategy	Simulation model	Crop	Reference
Data assimilation	Replacing the simulated fAPAR with satellite-derived fAPAR	MOSICAS	Sugarcane	Morel et al. (2014)
	Calibrating parameters based on satellite-derived LAI	WOFOST	Wheat	Huang et al. (2015)
	Initializing the model based on the emergence date from VI time series, then calibrating parameters based on satellite-derived Green Area Index	SAFY	Maize	Battude et al. (2016)
	Calibrating parameters at the county-scale based on aggregated LAI, then updating the simulation at the pixel-scale (30 m) based on satellite-derived LAI	SAFY	Maize	Kang and Özdoğan (2019)
Metamodels	Developing a simple linear regression between simulated yield and fAPAR	SUCROS	Sugarcane	Clevers et al. (1997)
	Developing a multiple linear regression between simulated yield and simulated GCVI, expanded with weather predictors	APSIM-Maize; APSIM-Soybean	Maize; Soybean	Lobell et al. (2015)
	Developing a Random Forest model between simulated yield and simulated GCVI, expanded with weather predictors	WheatGrow	Wheat	Zhao et al. (2023)
Hybrid models	Developing multiple machine learning models based on satellite fAPAR and simulated outputs (e.g., water-limited biomass, water storage at grain filling)	WOFOST	Barley; Potato; Wheat	Paudel et al. (2021)
	Developing a Random Forest model based on satellite GCVI and a simulated drought index	SALUS	Maize	Shuai & Basso (2022)
	Developing multiple machine learning models using drone-based VIs and simulated outputs (e.g., biomass, canopy cover)	AquaCrop	Wheat	Cheng et al. (2024)

Assimilation of spectral data into crop simulation models

The term "data assimilation" is often used interchangeably to describe different strategies to augment a crop simulation model with spectral data. Perhaps the most intuitive strategy is "forcing", which involves the direct replacement of simulated model state variables with those derived from spectral data (Fig. 2.4Aa, Bouman, 1995; Clevers et al., 2002; Morel et al., 2014; Yao et al., 2015). Due to their critical role in simulation models, the most commonly used are biophysical variables such as LAI or fAPAR. The primary advantage of this strategy is its simplicity, as it directly replaces the model subroutine, such as leaf area generation, with remotely-sensed data. However, its performance is highly dependent on the quality and frequency of remote sensing data. Errors in these estimates can propagate throughout the simulation, compromising its overall accuracy. Moreover, since simulation models typically operate at a daily time step while remotely-sensed data may have lower temporal resolution, interpolation introduces an additional source of error.

To mitigate some of these limitations, the “updating” strategy can be employed (Fig. 2.4Ab, Chen & Tao, 2020; Dhakar et al., 2022; Gaso et al., 2023; Kang & Özdoğan, 2019; Zare et al., 2024). In this approach, the model simulates state variables until a remotely-sensed estimate becomes available. At this point, the state variable is updated by a weighted combination of the simulated and remotely estimated values, and the simulation continues from this updated value until the next remotely sensed estimate becomes available. As the model is adjusted with each new remotely-sensed estimate, this approach is also referred to as “sequential assimilation”. Compared to forcing, a major advantage of updating is that the model is not entirely dependent on the quality and frequency of remote sensing estimates, as it continues the simulation even if such data are not available. A disadvantage, however, is the increased computational cost of updating the model through weighted adjustments with each new observation. To determine the weights between simulated and estimated values, the most commonly used algorithms include the Kalman filter, ensemble filter, particle filter, among others. For details on the advantages of each algorithm, the reader is referred to Huang et al. (2019).

Spectral data can also be employed in the “calibration” strategy to adjust the parameters of the simulation model (also referred to as “re-calibration” or “variational assimilation”, Fig. 2.4Ac, Battude et al., 2016; Gaso et al., 2021; Gilardelli et al., 2019; Huang et al., 2015, 2019). In this strategy, model parameters are iteratively adjusted to minimize the difference between simulated and remotely-sensed estimates, as defined by a cost function (e.g. root mean square error). In this approach, all remotely-sensed estimates are used simultaneously. For example, the maximum radiation use efficiency can be iteratively modified to find the value that minimizes the difference between simulated and remotely-sensed LAI (Gilardelli et al., 2019). The main advantage of this strategy is that it enables the development of a more realistic model by refining its parameters, leading to better simulation outcomes (Jin et al., 2018). However, it often requires multiple iterations to minimize the cost function, which increases the computational costs, and it relies on remote sensing data throughout the entire crop cycle, which may limit its feasibility for real-time applications (e.g., forecasting yield at mid-season).

A fourth strategy is the “initialization” of models, in which spectral data are used to estimate the input variables required to start running simulations (Fig. 2.4Ad, Battude et al., 2016; Han et al., 2019). Unlike previous strategies that focus on modifying state variables or parameters, this approach targets the variables required to initialize crop simulations. This is particularly useful for scaling models regionally, where there may be uncertainty regarding crop area and management practices (e.g., sowing date). For instance, by analyzing the temporal evolution of VIs, it is possible to estimate sowing or emergence dates to initialize the simulations (Han et al., 2019).

In this section, we provided a brief explanation of the different strategies for assimilating remotely-sensed spectral data into crop simulation models. For more details, the reader is referred to specific reviews (Dorigo et al., 2007; Jin et al., 2018; Huang et al., 2019).

Metamodels based on multiple crop simulations

This subgroup of models runs multiple simulations to calibrate simpler models ("metamodels") that can then be quickly applied to spectral data (Fig. 2.4B, Azzari et al., 2017; Clevers et al., 1997; Deines et al., 2021; Jin et al., 2017; Lobell et al., 2015; Zhao et al., 2023; Ziliani et al., 2022). The process is developed in four stages. First, a large number of simulations are conducted under realistic conditions of weather, soil, and management for the site of interest. From each simulation, the final yield and the time series of relevant variables—such as LAI, fAPAR, or other variables potentially derivable from spectral data—are recorded. In the second stage, these variables are translated into spectral data, using calibrated relationships (e.g., estimating a VI from simulated LAI, Nguy-Robertson et al., 2012). In the third stage, empirical models are calibrated between simulated yield and simulated spectral observations. The simplest approach involves fitting linear regressions between yield and VIs at specific stages, although these models can be extended to include weather predictors (Lobell et al., 2015) or to leverage machine learning algorithms (Zhao et al., 2023). Finally, in the fourth stage, the calibrated models are applied to real spectral data to predict yield.

An example of this subgroup of models is the system developed by Lobell et al. (2015), called the "Scalable Satellite-based Crop Yield Mapper". Based on APSIM simulations, linear regression models were calibrated between simulated maize and soybean yields and simulated observations of Green Chlorophyll VI (GCVI, estimated from simulated LAI) and weather variables. These regression models were then applied to satellite imagery, allowing the mapping of crop yields at a spatial resolution of 30 m across large areas without requiring field data for model calibration. Over time, this system was improved and replicated for other regions and crops (Azzari et al., 2017; Deines et al., 2021; Jin et al., 2017). For instance, Deines et al. (2021) evaluated predictive models for maize with yield monitor records (2008–2018). Their models achieved an R^2 of 0.4 at the subfield scale, which increased to 0.45 at the field scale and 0.69 at the county scale. Similarly, the RMSE decreased from 2.45 to 1.85 and 1.25 Mg ha^{-1}, respectively.

Hybrid models with crop simulation outputs and spectral data as individual predictive variables

This subgroup consists on calibrating empirical models using both simulation outputs and spectral data as individual predictors. The term "hybrid models" arises from calibrating data-driven models with the outputs of process-based models (Fig. 2.4C, Gaso et al., 2024; Li et al., 2023; Paudel et al., 2021; Shuai & Basso, 2022; Zhuang et al., 2024). There are some reasons for preferring hybrid models over the previous subgroups. For example, these models are expected to outperform data assimilation when the derivation of state variables from spectral information contains errors or biases, as assimilation of such data would only propagate these uncertainties (e.g., consistent overestimation of LAI). In addition, spectral data can provide valuable information for yield prediction that is difficult or impossible to assimilate into crop simulation models due to model architecture constraints (e.g. specific reflectance values). Similarly, hybrid models are supposed to outperform metamodels when the crop simulation model does not accurately capture the relationship between yield and biophysical variables (e.g. yield vs. LAI). To address these limitations, this subgroup of models treats simulation outputs - such as biomass, canopy cover or even yield - and spectral data as separate predictors and incorporates their potential errors into the overall error of the empirical model.

There are successful examples of this methodology. Feng et al. (2020) calibrated Random Forest models to predict wheat yield at the plot scale across 29 sites in Australia (2008–2019), using simulated biomass by APSIM, NDVI from MODIS satellite, and weather variables, including frost and heatwaves. They found that model accuracy improved as the crop cycle progressed, reaching a final error of 17% at physiological maturity, with the most influential variables being those simulated by APSIM. In this case, NDVI had a marginal contribution, likely due to the relatively low spatial resolution of MODIS compared to the experimental plots. Similarly, Cheng et al. (2024) predicted yield in mulched wheat plots in China (2017–2021). The authors calibrated machine learning models using predictors that combined simulated outputs from AquaCrop—such as biomass and canopy cover—with VIs obtained from drones. This approach proved more accurate than both AquaCrop simulations and models based solely on spectral data, achieving the lowest error during heading (8%) compared to earlier stages, such as turning green and jointing (17% and 13%, respectively).

Discussion

In the previous section, we identified and described crop yield prediction frameworks and highlighted some of their strengths and limitations. Now, we discuss three key aspects common to all of them: (1) the data required for model calibration, (2) the trade-off between interpretability and predictive power, and (3) their spatio-temporal transferability (Table 2.4). Additionally, we discuss some potential pathways to improve crop yield prediction models.

Table 2.4: Summary of each framework according to the data required for model calibration, accuracy, interpretability, and spatio-temporal transferability.

Group	Subgroup	Requires translation from spectral to biophysical data?	Requires other data sources?	Requires ground-truth yield data for calibration?	How effectively does it capture yield variability?	How interpretable is the model?	How easy is to anticipate transferability?
Based directly on spectral data	Simple statistical models	No	Only those that improve accuracy	Yes. It does not need larger datasets	Predictors must be linearly related to yield	Interpretable due to its simple algorithm	Easier due to high interpretability
	Machine learning models	No	Only those that improve accuracy	Yes. It needs larger datasets to train models (e.g., thousands or more data points)	Handles raw predictive variables with complex, non-linear relationships with yield	May be complex due to non-linear relationships and large number of predictors	Complex due to reduced interpretability
Based on derived biophysical variables	Empirical models	Yes	Only those that improve accuracy	Yes. Varies with the algorithm (statistical or machine learning models)	Varies with the algorithm (statistical or machine learning models)	More interpretable due to biophysical variables; Simple if statistical; Complex if machine learning	Varies with the algorithm, easier if statistical; more complex if machine learning
	Resource-based models	Yes	Resource data (e.g., incident radiation or available soil water)	Yes. It does not need larger datasets	Resource capture and use efficiencies must explain most of the yield variability	Interpretable due to its process-based structure	Easier due to high interpretability
Integrated with crop simulation models	Data assimilation	Yes	Weather, soil and management data to run simulations. Simulations due to low-quality inputs may be corrected with assimilated data	Not needed if the simulation model was previously validated	Spectral data must better represent reality than crop simulations	Interpretable due to its process-based structure	Easier due to high interpretability
	Metamodels	Yes	High-quality weather, soil and management data to run simulations	Not needed if the simulation model was previously validated	Simulations must correctly reflect the relationship between yield and predictor variables (statistical or machine learning models)	More interpretable due to its process-based nature; Simple if statistical; Complex if machine learning	Varies with the algorithm, easier if statistical; more complex if machine learning
	Hybrid models	Not necessary	High-quality weather, soil and management data to run simulations	Yes. Varies with the algorithm (statistical or machine learning models)	Spectral data and simulated outputs must capture different portions of yield variability (statistical or machine learning models)	Simple if statistical; Complex if machine learning	Varies with the algorithm, easier if statistical; more complex if machine learning

Data required for model calibration

For frameworks that rely on biophysical variables, the first step is to translate spectral data into these variables using an appropriate model. A robust but labor-intensive approach is to develop or validate models based on local crop measurements (Huang et al., 2015; Zare et al., 2024). However, for large spatial scale studies, it is often more practical to adopt a previously published model (Lobell et al., 2015). While this approach eliminates the need for field sampling, it introduces uncertainty, as the original model may have been developed under different environmental, management or crops conditions (Colombo et al., 2003; Gitelson et al., 2014), or may be influenced by the interaction between the sensor and site conditions where it was calibrated (Epiphanio & Huete, 1995). Currently, satellite-derived products provide periodic estimates of biophysical variables (Johnson et al., 2015; López-Lozano et al., 2015). In addition, generic methodologies, such as the one proposed by Sellers et al. (1996) for the estimation of fAPAR, provide alternative solutions (Campoy et al., 2023).

Requirements of data volume (quantity) and diversity (different data sources) can limit the available frameworks for yield prediction. Machine learning models need large datasets for proper calibration (Kamir et al., 2020), but when data are limited, incorporating prior knowledge by selecting predictors with a causal link to yield can be a feasible alternative (Maestrini et al., 2022; Rudin, 2019). Beyond spectral data, other data sources also play a crucial role, particularly in crop simulation models. These simulations rely on detailed weather, soil, and agronomic management data, and the lack of local data or low quality of these inputs can limit their practical application (Rosenzweig et al., 2013). This is particularly true for frameworks involving metamodel development or hybrid models, as poor simulations reduce their usefulness. In contrast, data assimilation frameworks may be less dependent on these inputs, as spectral data can correct errors in simulated outputs caused by inaccuracies in other inputs, provided the spectral data are reliable. On the other hand, other frameworks that are not integrated with simulation models do not strictly require these additional data sources, and their inclusion is only justified if they contribute to reducing prediction errors (e.g., Cai et al., 2019).

Trade-off between Interpretability and predictive power

A first factor affecting this trade-off is the choice of predictor variables. Comparatively, crop yield prediction models based directly on spectral data are less interpretable than those based on biophysical variables. This is particularly true for people unfamiliar with remote sensing concepts, for whom a VI value may seem abstract, whereas a biophysical variable such as LAI is conceptually more intuitive. In terms of predictive power, there is no reason to expect differences between the use of spectral data or biophysical variables, unless the translation requires more than a simple regression model.

A second factor affecting this trade-off is the algorithm. Linear regressions between yield and a few predictor variables, whether VIs or biophysical indicators, yield highly interpretable models. However, achieving predictive power requires careful construction of variables that have a good linear relationship with yield. Similarly, frameworks based on resource capture and use efficiencies, as well as those integrated with crop simulation models, offer high interpretability since these models synthesize much of the current knowledge about crop processes, representing a more functional biological system. However, their predefined structure constrains the patterns they can capture, limiting their predictive power. For example, crop simulation models updated with spectral data can suffer from "phenological shift", a phenomenon in which the model correctly replicates the evolution of the updated state variable but fails to adjust for the crop phenology. Since phenology controls key sub-processes involved with yield formation, such as biomass partitioning to harvestable organs (harvest index), this misalignment could substantially increase prediction error (Curnel et al., 2011; Yang et al., 2023).

Machine learning models are effective in overcoming some of these limitations. These algorithms can capture non-linear effects and interactions between predictors, such as the differential effect of VIs on yield depending the phenological stage or temperature exceeding a certain threshold (Feng et al., 2020; Zhang et al., 2019). As a result, these crop yield prediction models often achieve higher predictive power than linear or simulation models, even when trained on raw predictor variables (Cai et al., 2019; Gaso et al., 2024; Ma et al., 2024; Schwalbert et al., 2020). However, their agronomic interpretability may be limited, making it difficult to understand how predictive variables contribute to yield prediction in terms of direction and magnitude.

Spatio-temporal transferability of the models

The primary goal of yield prediction models is to maintain accuracy when applied to new growing conditions, such as different years or regions. To achieve this, it is crucial to improve our ability to anticipate when they will perform well, and more interpretable models make this easier. For example, consider a linear regression model that predicts maize yield based on absorbed radiation throughout the crop cycle (as in Gallo et al., 1985), and supposed it was calibrated using data from years without heat stress and then transferred to predict in a year with heat stress. It is likely that the model would overestimate yield due to the issue that heat stress reduces harvest index without significantly affecting absorbed radiation (Cicchino et al., 2013). In more complex models, predictions depend on a larger number of variables and intricate relationships. As a result, anticipating how well they will transfer to new conditions becomes more challenging.

Regardless of the crop yield prediction framework, it is essential to evaluate models under conditions that accurately represent their intended application (Filippi et al., 2025). As yield prediction models are almost always designed to predict in new years, it should be mandatory to assess their temporal transferability. This process is commonly implemented through the "Leave-One-Year-Out" procedure (e.g. Bolton and Friedl, 2013; Cai et al., 2019; Schwalbert et al., 2020). However, if the final goal is to forecast into future years, an even more realistic evaluation is to simulate operational conditions, in which the model is calibrated only with previous years (e.g., Li et al., 2023; Ma et al., 2021; Paudel et al., 2021). In other cases, the aim may be to use data from one region to make predictions in another, in which case it is essential to assess spatial transferability (e.g. Becker-Reshef et al., 2010; Ma et al., 2023; Priyitankto et al., 2023).

Some pathways for enhancing crop yield prediction models

One way to improve these models is to incorporate additional data sources beyond traditional spectral reflectance. Some relatively unexplored sources have the potential to enhance predictive power: (1) radar data from remote sensors, particularly valuable in regions with frequent cloud cover where high-quality optical imagery is scarce. However, the main challenge is pre-processing to reduce noise levels, although recent advances have improved their applicability and translation into biophysical variables (Mullisa et al., 2021; Veloso et al., 2017); (2) VIs that assess photosynthetic efficiency, beyond conventional VIs focused on absorbed radiation (e.g., NDVI or EVI for fAPAR or LAI estimation). Among these, solar-induced chlorophyll fluorescence (SIF) may provide insights into radiation use efficiency for biomass production, as it originates from chlorophyll radiation emission under stress (Mohammed et al., 2019; Zeng et al., 2022). Other indices based on radiation reflectance such as the photochemical reflectance index (PRI), also estimate this efficiency, although their dependence on hyperspectral data currently limits their use to private satellites or hand-held and drone-mounted devices (Garbulsky et al., 2011).

Another way to improve models is to make them more interpretable, which increases their universality and practical applicability. One approach is to incorporate predictive variables with a stronger causal relationship to yield. The current trend prioritizes accuracy, often leading to the inclusion of multiple predictors, diverse data sources, and complex algorithms. However, these models may be difficult to communicate, limiting their adoption in scenarios where prediction error is only one of several considerations (Rudin, 2019). In this context, integrating prior knowledge into predictor variables can help balance accuracy and interpretability. For example, instead of using a VI time series directly, differentiating it by key phenological stages, such as critical crop periods, can guide the algorithm, reduce prediction error, and clarify how the prediction is constructed (Pei et al., 2025). As a result, more interpretable models function not only as predictive tools, but also as diagnostic tools to support crop management decisions.

Summary

In this study, we revisited existing frameworks for crop yield prediction based on remotely sensed spectral data (Fig. 2.1). Since the first efforts in the 1970s, multiple methodologies have emerged to integrate this information into predictive models. Here, we identified and synthesized three major groups of frameworks, each with its subgroups: (1) models based directly on spectral data, including simple statistical and machine learning models (Fig. 2.2, Table 2.1); (2) models based on derived biophysical variables, classified as empirical or based on resource capture and use efficiencies (Fig. 2.3, Table 2.2); and (3) models integrating spectral data with crop simulation models, including data assimilation, metamodels, and hybrid models (Fig. 2.4, Table 2.3). We then discussed key aspects common to all approaches, such as data requirements for model calibration, interpretability, accuracy, and spatio-temporal transferability. We found considerable heterogeneity among frameworks (Table 2.4), making it difficult to define a single best approach.

Instead, the optimal choice depends on the objective, available data, and expertise. Future research should focus on the development of crop yield prediction models with high spatio-temporal coverage and resolution, especially at the field or sub-field scale, as such models remain scarcely explored in the scientific literature. Advancing this research efforts at field or sub-field level would provide a clear pathway to improving the practical relevance of crop yield prediction models, ultimately improving farmers' decisions and livelihoods worldwide.

3. Spatio-temporal transferability of satellite-based wheat yield prediction models

Abstract

Developing universal crop yield prediction models that are transferable across regions and years is crucial for food security. Here, we aimed to develop satellite-based wheat yield prediction models that only require location and sowing date as user inputs, and assess their accuracy when transferred to unknown years or regions. We analyzed 5,265 wheat fields across seven regions in Argentina (2018–2023), including yield, location, sowing date, and anthesis date from a phenological model. We integrated absorbed radiation derived from satellite imagery with weather data to develop Elastic-Net models predicting yield from 50 days before anthesis to 20 days after, using three frameworks: (1) training and evaluating within the same year/region, (2) evaluating in unknown years/regions to test transferability, and (3) evaluating in future years with cumulative training from previous years. Absorbed radiation was the key predictor, whose importance increased as anthesis approached, while weather variables were marginally important. Yield prediction models lost more accuracy when predicting unknown regions rather than years, with spatial transfer errors up to three times greater than temporal transfer. When predicting future years, the median error at anthesis was 0.67 t ha^{-1} (17%). This study demonstrates the robustness of satellite-based models to predict wheat yield in unknown years, while highlighting the challenges of transferring these models to unknown regions. Since models were driven by absorbed radiation, the remaining error likely arises from differences in conversion efficiency into yield, varying more across regions than years. These field-scale models combine accuracy, pre-harvest anticipation and simplicity, making them valuable tools for various stakeholders.

Keywords: estimation; forecast; machine learning; NDVI; fAPAR; Sentinel-2; MODIS

Introduction

Global food demand is growing faster than production, threatening consistent access to sufficient, nutritious, and safe food (Lal, 2016). Addressing this issue requires optimizing the entire food production chain, with a focus on advanced crop yield prediction models (Clark et al., 2023; Schauberger et al., 2020). These models provide real-time forecasts throughout the growing season, aiding in the reduction of price volatility, the identification of affected regions, and the effective allocation of resources. Timely and accurate data on global food production is crucial for food security (Burke et al., 2020; Fritz et al., 2019).

Ideally, crop yield prediction models should meet specific criteria to be practical. First, models should provide early-season yield predictions to support tactical decisions, enabling field-scale optimization for farmers and regional-scale resource allocation for governments and NGOs (Paveley et al., 1997; van der Velde et al., 2019). Second, models should require minimal input data. Currently, most "low-input" models operate at regional scales, while field or sub-field models often require extensive inputs or rely on datasets with limited spatial or temporal coverage, thereby limiting their adoption (Basso & Liu, 2019). Third, agronomically interpretable models are desirable to serve both as a predictive tool and resource for post-season analysis, enabling stakeholders to assess management and environmental factors affecting yields (Hu et al., 2023; Lobell, 2013).

The main challenge for yield prediction models is to maintain accuracy when applied to unknown years or regions (temporal and spatial transfer, respectively). Temporal transfer is critical since the main goal of these models is to forecast yield into future, which can be tested by training models exclusively on data from previous years. Spatial transfer, on the other hand, assesses the applicability of models to regions without historical data, which can be tested by training models in specific regions and applying them directly elsewhere. According to the meta-analysis by Schaurberger et al. (2020), only half of yield prediction models included independent evaluations using data outside the training set. Furthermore, even among those that did, many relied on random data splits such as the classical 70% for training and 30% for evaluation (Filippi et al., 2025). Such evaluations are not entirely spatio-temporal independent and tend to produce artificially low errors, thereby limiting our ability to anticipate model performance in real-world applications (Morales & Villalobos, 2023). Nevertheless, some successful examples exist: Cai et al. (2019) utilized the "Leave-One-Year-Out" method to evaluate temporal transferability of wheat yield prediction models in Australia at the county scale, achieving consistent accuracy with the exception of a dry year, while Priyatikanto et al. (2023) applied "transfer learning", a technique rooted in deep learning, to predict maize yield at the county scale across regions and years in United States, finding spatial transfer more challenging due to higher variability between regions than years. Despite advances made in this field, studies utilizing large, multi-year, multi-regional datasets and field-scale models that are agronomically interpretable remain scarce.

In this context, satellite data help to develop such models because vegetation indices estimate the fraction of photosynthetically active radiation absorbed by crops (fAPARgreen; Sellers et al., 1987), thereby facilitating the estimation of crop growth rate (Hiker et al., 2008). Many studies have modeled yield using satellite data from single or multiple crop phenological stages (Bolton & Friedl, 2013; Lobell et al., 2015; Schwalbert et al., 2020). As satellites capture absorbed radiation, most of the model error arises from variability in the conversion efficiency of absorbed radiation into yield, driven by environmental and crop-specific factors. Consequently, the transferability of remotely-sensed models will depend on the degree of variability in this efficiency between regions and years. To ensure that these empirical models can reliably predict yield, it is critical to assess their spatio-temporal transferability.

Wheat (*Triticum aestivum L.*) is the most widely cultivated crops globally and is key to ensuring food security (Shewry & Hey, 2015). Due to the heterogeneity of production environments, the development of yield prediction models that are transferable across years and regions remains a significant challenge. As noted above, satellites capture the radiation absorbed by the crop, but not the efficiency with which it is converted into grain yield. Temporal and spatial variations on this efficiency in wheat may be due to environmental (Barlow et al., 2015; Rodriguez & Sadras, 2007), genetic (Miralles & Slafer, 1997; Reynolds et al., 2012), or management factors (Caviglia & Sadras, 2001; Estrada-Campuzano et al., 2012), ultimately hindering model transferability.

The objectives of this study were to: 1) develop satellite-based wheat yield prediction models at the field scale using only location and sowing date as user inputs, and assess their accuracy when trained and evaluated within the same year or region, 2) assess the loss of accuracy when transferring models to predict yield in unknown years or regions, and 3) assess a specific case of temporal transfer by simulating operational conditions (e.g. predicting future years from the cumulative training from previous years).

Methodology

General description of the methodology

The conceptual framework presents the three main stages on this study: 1) compilation of an agronomic database, 2) compilation of satellite and weather databases, and 3) modeling for yield prediction. Briefly, for the first stage, we compiled an agronomic database with yield, location, sowing date, and genotype for 5,265 wheat fields across six years and seven regions in Argentina.

We estimated crop phenology, mainly anthesis dates, using a phenological model (Fig. 3.1, stage 1). For the second stage, we retrieved time series of fAPARgreen from Sentinel-2 and MODIS satellites and weather variables from NasaPower (Fig. 3.1, stage 2). Lastly, for the third stage, we tested different yield models to explore the three objectives. For objective 1, we trained and evaluated yield prediction models within the same year and region using data from 50 days before to 20 days after anthesis, at 10-day intervals. For objective 2, we assessed the loss of accuracy of models when transferred to predict unknown years and regions, not included in the training dataset. Lastly, for Objective 3, we assessed a specific case of temporal transfer by simulating operational conditions and predicting future years using training data from previous years only (Fig. 3.1, stage 3).

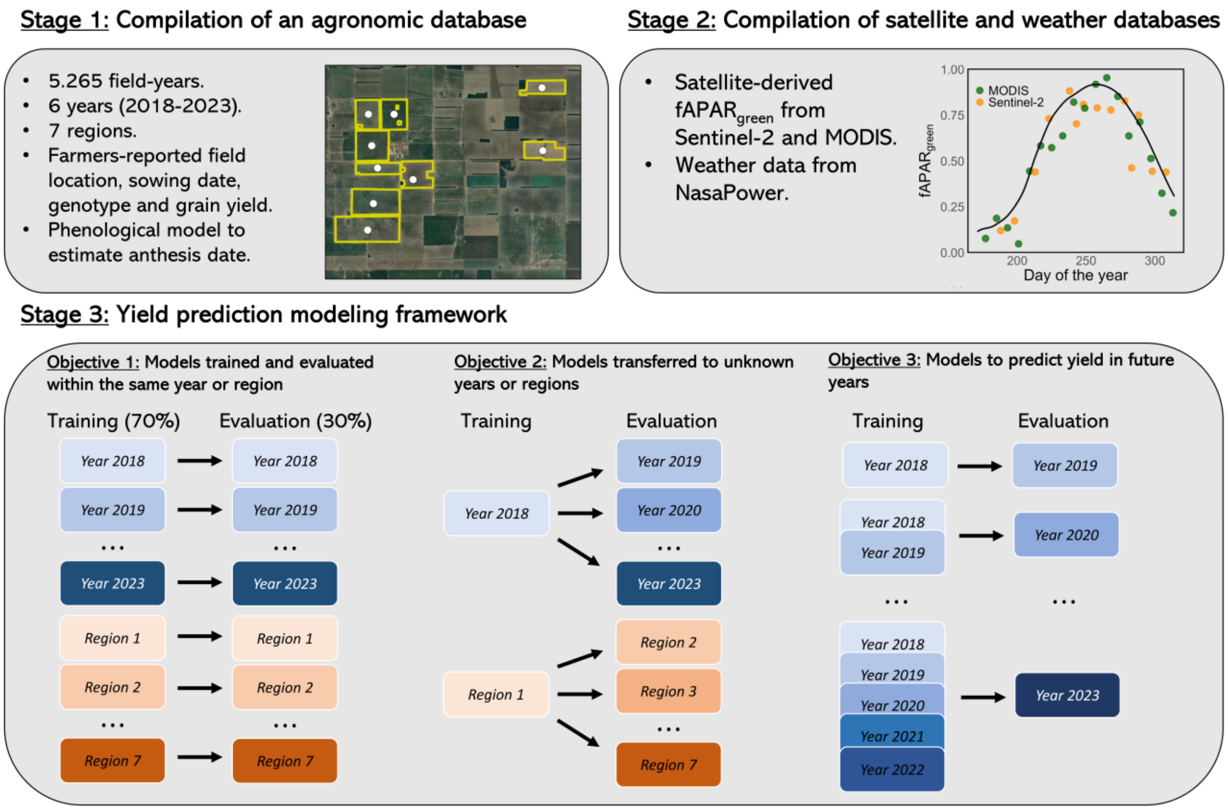

Figure 3.1: *Conceptual framework of the study. Stage 1: compilation of an agronomic database with field-year information and estimated anthesis dates. Stage 2: compilation of fAPARgreen time series derived from Sentinel-2 and MODIS satellites and weather data from NasaPower. Stage 3: yield prediction modeling framework: training and evaluation within the same year or region (Objective 1), evaluation in unknown years or regions (Objective 2), evaluation in future years (Objective 3).*

Study region

The entire study region spans from 26°S to 39°S and 57°W to 65°W in Argentina, encompassing diverse environmental conditions and crop management practices (Fig. 3.2A and 3.2B). We partitioned the entire region into seven regions defined by the Consortium of Regional Experimentation of Agriculture organization (CREA), a non-profit organization that connects more than 2,000 farming companies organized into sub-regional groups. These regions are heterogeneous in environmental and management conditions (Calviño and Monzon, 2009; Magrin et al., 2005): frost-free period ranges from 140 days in the south to 180 days in the north, mean annual temperatures from 14°C to 22°C, and annual precipitation from 500 mm in the southwest to 1,200 mm in the east. The dominant soils are mollisols and vertisols. Individual fields are extensive, with a median size of 57 ha in our dataset (5th percentile 14 ha, 95th percentile 184 ha). Almost all fields are no-till and spring wheat is sown from April to August. As the sowing date progresses, farmers switch from long to short cycle genotypes to time anthesis just after the last frost of the season. Thus, anthesis dates converge for different sowing dates and genotype cycles, but vary significantly with latitude. In the north (latitude < 30°S) anthesis occurs in mid-September, while in the south (latitude > 36°S) it occurs in mid-November. Harvesting takes place about 50 days after anthesis.

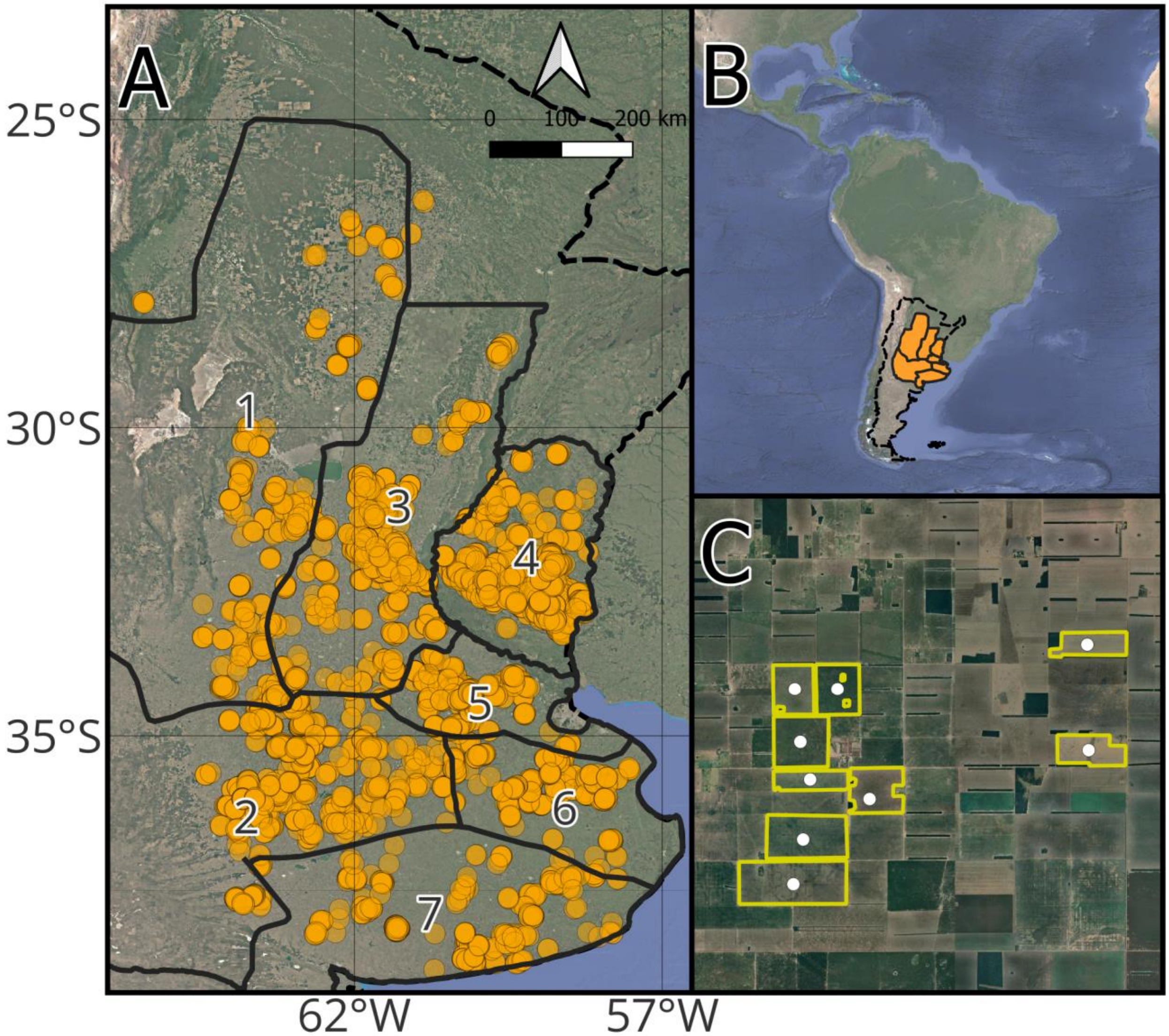

Figure 3.2: *Study region and field boundary delineation example. A) Entire study region: orange dots refer to the 5,265 field-year observations, solid black lines with numbers refer to the seven regions, and dashed black lines refer to Argentina's national boundaries. Field-years outside regions were assigned to the nearest one. B) Location in the continent and regions. C) Field delineation: white dots refer to farmer-reported geographic coordinates, and yellow polygons refer to delineated field boundaries.*

Agronomic dataset and phenology estimation

We compiled a farmers database on wheat fields spanning from 2018 to 2023 from CREA Traced Agricultural Data (DATCREA, https://www.crea.org.ar/dat-crea/), comprising 6,724 field-year records with data on year, coordinates, area, genotype, sowing date, and grain yield. For each field-year, we delineated boundary polygons using farmer-provided coordinates and verified against Sentinel-2 satellite images from each year to ensure alignment with reported areas (Fig. 3.2C, QGIS.org, 2022). A total of 1,007 records were discarded due to inaccurate coordinates or ambiguous field boundaries.

To estimate crop phenology, we developed a model to predict emergence, anthesis, and physiological maturity dates based on sowing date, latitude and longitude. We utilized the web-based model “CronoTrigo” designed for wheat in Argentina (Miralles et al., 2007, available in Spanish version: http://cronotrigo.agro.uba.ar/index.php/). CronoTrigo is freely available after registration and integrates historical climate records for 256 counties in the Argentine wheat region and 76 commercial genotypes calibrated for photoperiod and vernalization responses. CronoTrigo requires inputs of genotype, county, and sowing date and provides estimated dates for emergence, anthesis, and physiological maturity. For this study, we estimated phenological dates for 4,533 field-years in our dataset that grew genotypes available in CronoTrigo. We then regressed these phenological dates against sowing date, latitude, and longitude and obtained a model to estimate phenological dates for any field-year without requiring specific genotype information, relying solely on field sowing date and coordinates. The mean absolute errors were 0.81 days for emergence, 4.19 days for anthesis, and 3.5 days for physiological maturity (Fig. S1-S3).

Satellite and weather datasets

For each field-year, we calculated NDVI from March 1 to December 31 for all pixels using Sentinel-2 (COPERNICUS/S2_SR_HARMONIZED, bands 4 and 8) and MODIS (MODIS/061/MOD09Q1, bands 1 and 2) satellite imagery in Google Earth Engine (Gorelick et al., 2017). Sentinel-2 has a spatial resolution of 10 m for NDVI bands, a temporal resolution of 5–6 days, and high-quality images in our study region from 2019 onwards. MODIS, operational for over 20 years, has a spatial resolution of ~250 m with a daily temporal resolution; however, the product used provides the highest-quality pixel reflectance every 8 days. To avoid including pixels overlapping extra-field areas, we applied a negative buffer of 20 m to Sentinel-2 images and 100 m to MODIS images around each polygon. For Sentinel-2, we applied two filters to the data: 1) selecting images with less than 50% cloud cover across the mosaic, and 2) retaining only pixels classified as non-cloud and non-cirrus, based on the 'QA60' band. We discarded a total of 240 field-years due to satellite data gaps exceeding 16 days.

Based on the NDVI. for each field-year we calculated the daily time series of the fraction of incident photosynthetically active radiation absorbed by the crop (fAPARgreen, Fig. 3.3), transforming each pixel into fAPARgreen using models calibrated for wheat by Pellegrini et al. (2020), specifically for Sentinel-2 and MODIS in our study region. Then, we averaged fAPARgreen values from all pixels of each field-year and date. In order to produce daily data and to reduce the residual noise from clouds and cirrus, we applied a two-step process to smooth the fAPARgreen time series. In the first step, we fitted a local regression and computed the differences between predicted and observed fAPARgreen values. If the predicted value exceeded the observed value, we kept the positive difference; otherwise, we set the difference to zero. In the second step, we fitted a new local regression, this time weighted by the difference vector. This weighting reduced the influence of observations where fAPARgreen decreased, likely due to errors from clouds or aerosols rather than natural crop development (Cai et al., 2017). Local regressions were implemented using the loess() function in R software (R Core Team, 2023), setting both regressions to degree two with a span of 0.2.

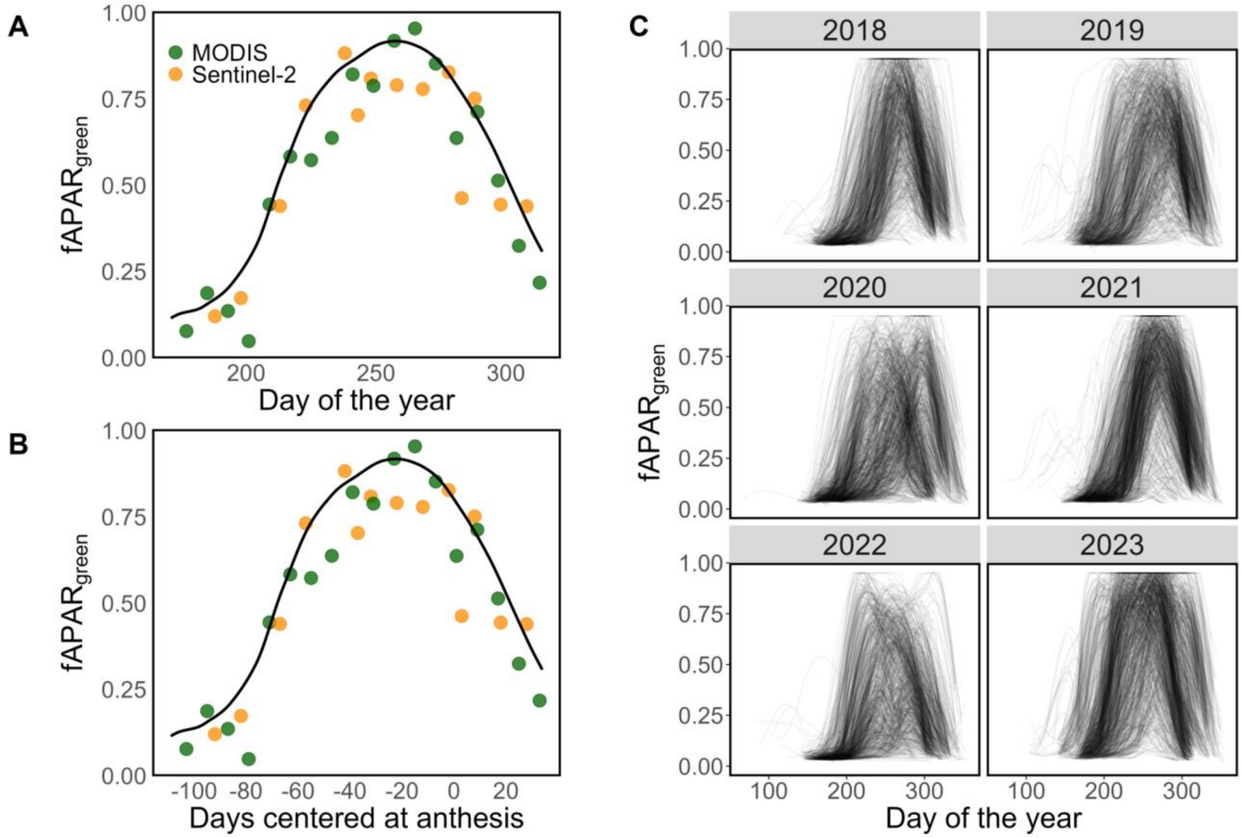

Figure 3.3: *fAPARgreen estimates and transformations. A) Example of a time series of fAPARgreen for a field-year. Each point refers to an estimate of fAPARgreen from MODIS NDVI (green points) and Sentinel-2 NDVI (orange points) using the models of Pellegrini et al. (2020). The black solid line refers to the final smoothed fAPARgreen time series. B) Same as A, but expressed relative to days centered at anthesis, with phenology estimated based on location and sowing date. C) Smoothed fAPARgreen time series for all field-years. Each black line refers to a field-year.*

For each field-year, we compiled a daily weather database. We retrieved daily data from the Prediction of Worldwide Energy Resource, which reports variables at a spatial resolution of ~60 km (NASA POWER, https://power.larc.nasa.gov/data-access-viewer/): incident photosynthetically active radiation, precipitation, and mean, minimum, and maximum air temperature at 2 m, as well as dew point temperature. We calculated the daily vapor pressure deficit for each field-year from daily maximum, minimum, and dew point temperature (FAO, https://www.fao.org/4/x0490e/x0490e07.htm).

Data analysis

Before analyzing the data, we removed 212 field-years from the database, 118 of them because fAPARgreen values were above 0.25 at sowing, indicating likely incorrect sowing dates (examples in Fig. S4). The 94 additional field-years were excluded because the radiation-to-grain conversion efficiencies (calculated as grain yield divided by total absorbed radiation) exceeded twice the theoretical maximum of 1.5 g MJ^{-1}, based on a maximum radiation use efficiency for biomass for spring wheat of 3 g MJ^{-1} (Reynolds et al. 2012) and harvest index of 0.5 (Foulkes et al. 2011). Those fields were likely to be summer crops rather than wheat (examples in Fig. S5). The final dataset included 5,265 field-years.

For the first objective (assess model accuracy when trained and evaluated within the same year or region), we randomly split field-years within each year and region to train individual models specific to each year and region (70% training, 30% evaluation). Prediction times ranged from 50 days before anthesis to 20 days after, at 10-day intervals. To ensure robust results, we repeated the process of splitting, training, and evaluating the model 100 times per year and region, using the median error as the model evaluation metric. In addition, we applied the “Leave-One-Covariate-Out” method to quantify the importance of each predictor. For each trained model for a year or region in a given prediction time, we individually removed each predictor variable and recalculated error metrics, obtaining a relative error increase. In order to assess the overall importance of each predictor variable at a given prediction time, we averaged the relative error increase across years and regions.

For the second objective (assess the loss of accuracy when transferring models to predict unknown years or regions), we developed yield models to predict each year or region based on training data on all possible combinations of the remaining years or regions. For testing temporal transferability, each of the six years (2018-2023) was predicted using 31 different models, with training data ranging from one to five years. Similarly, to assess spatial transferability, each of the seven regions was predicted using 63 different models, with training data ranging from one to six regions. We calculated the error metrics for each temporal and spatial evaluation, and compared them to those in the previous objective to assess the loss of accuracy. We limited this objective to the prediction time with the lowest error identified in the previous objective.

For the third objective (simulate operational conditions), we predicted an entire year using models trained exclusively on data from prior years. For instance, models for 2019 were trained with 2018 data, while models for 2023 were trained with data from 2018 to 2022. Since 2018 was the first year in our dataset, we did not predict this initial condition.

For all models, we used five predictor variables: fAPARgreen, incident PAR (IPAR), precipitation (Prec.), temperature (Temp.), and vapor pressure deficit (VPD). We incorporated fAPARgreen differently depending on the prediction time: before and in anthesis, we included fAPARgreen with its value at the prediction time, while for predictions after anthesis, we included the cumulative daily integral of fAPARgreen from anthesis. Weather variables were summarized from the emergence date to the prediction time as: cumulative incident photosynthetically active radiation (MJ m^{-2}), cumulative precipitation (mm), average daily temperature (°C), and average daily vapor pressure deficit (MPa).

We employed the Elastic-Net algorithm, a linear regression method that combines Lasso and Ridge penalties to minimize the sum of squared errors, select relevant variables, and reduce overfitting (Zou and Hastie, 2005). The model structure is defined in Eq. 3.1:

$$Yield = \beta_0 + \beta_1 * \mathrm{fAPAR_{green}} + \beta_2 * IPAR + \beta_3 * \mathrm{Prec.} + \beta_4 * Temp. + \beta_5 * VPD + Error \quad (Eq. 3.1)$$

where β_0 is the intercept of the model, and β_1 to β_5 are the slopes for each predictor variable. If a variable is not important, its slope takes a value equal to or close to 0 (Zou and Hastie, 2005). We optimized the hyperparameters through 5-fold cross-validation on the training dataset, testing different combinations of α, the Lasso-Ridge balance, and λ, the regularization intensity (details in Appendix A). We selected the best hyperparameter combination on the training dataset and then used it to predict field-years for the evaluation dataset. We assessed model performance using two metrics: median absolute error (MAE, t ha^{-1}), and median absolute percentage error (MAPE, %), defined in Eq. 3.2 and Eq. 3.3:

$$MAE\ (t\ ha^{-1}) = Median(\,|\,Predicted\ yield - Observed\ yield\,|\,) \quad (Eq. 3.2)$$

$$MAPE\ (\%) = Median\left(\frac{|\,Predicted\ yield - Observed\ yield\,|}{Observed\ yield}\right) \quad (Eq. 3.3)$$

All the statistical analysis was conducted in R software (R Core Team, 2023), using the libraries tidyverse for data wrangling and visualization (Wickham et al. 2019), and glmnet() for fitting Elastic-Net models (Friedman et al. 2021).

Results

Models to predict yield within the same region or year

When predicting yield in the same region or year as the training dataset, errors decreased as the crop cycle progressed, with variations observed between regions and years (Fig. 3.4). Overall, the

MAE decreased from 0.83 t ha^{-1} at 50 days before anthesis to 0.61 t ha^{-1} at anthesis, while the MAPE decreased from 21% to 16%. After anthesis, errors stabilized at similarly low levels. In some years or regions, errors decreased significantly over the crop cycle (e.g., region 1 or year 2020), while in others, the reduction was more gradual (e.g., region 3 or year 2021). In a few cases, errors remained consistent throughout the crop cycle (e.g., region 4, years 2018 and 2019).

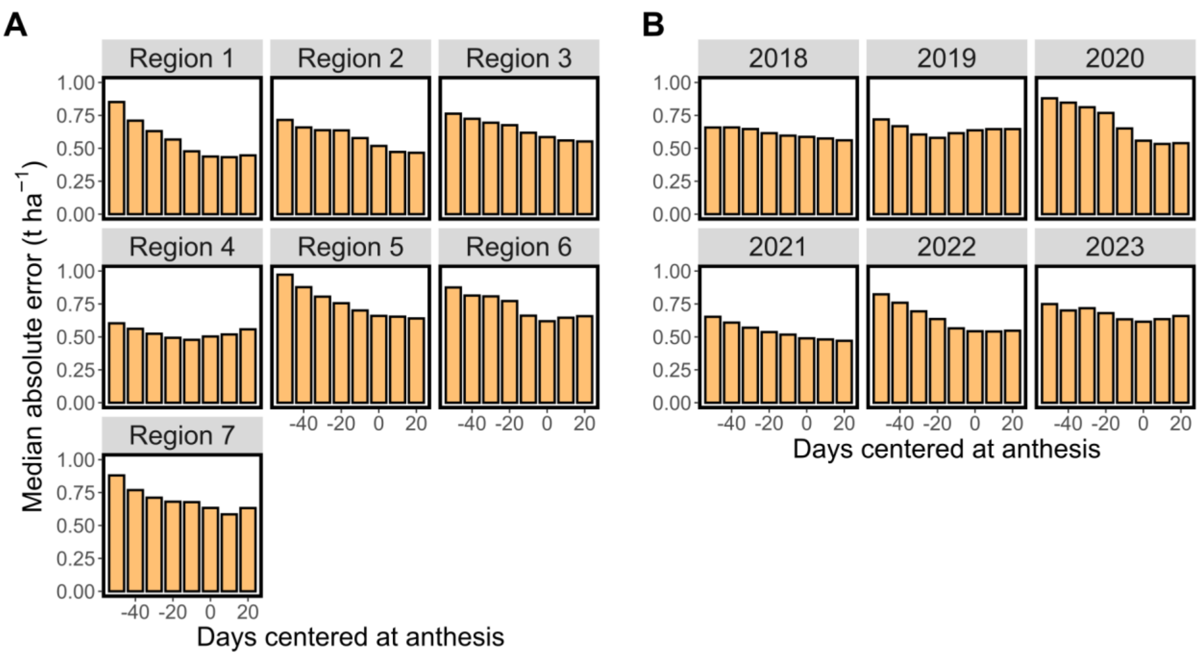

Figure 3.4: *Median absolute errors (MAE, t ha^{-1}) as a function of the days centered at anthesis. Each panel represents a predicted region (A) or year (B) with a model trained on data from the same year or region.*

The fraction of incident photosynthetically active radiation absorbed by the crop (fAPARgreen) was the most important variable for yield prediction, with its importance increasing as the crop cycle progressed (Table 3.1). Removing fAPARgreen from the model consistently increased the MAE, from 16% at 50 days before anthesis to 63% at anthesis, maintaining similar levels after anthesis. This latter scenario occurs since fAPARgreen at anthesis and its cumulative integral after anthesis are highly correlated (r = 0.98), conveying nearly identical information. In contrast, weather variables had a minimal effect on MAE, with changes always below 6%. Models trained on the complete dataset are shown in Table S1.

Table 3.1: Relative increase in median absolute error (MAE) after excluding each predictor variable from the model, compared to the full model with all variables included (%). Values represent the average relative increase in MAE across regions and years for each prediction time, expressed as days centered at anthesis.

Removed variable	**Days centered at anthesis**							
	-50	-40	-30	-20	-10	0	10	20
$fAPAR_{green}$	16	25	33	42	51	63	67	64
Incident PAR	2	2	2	1	2	1	1	1
Temperature	5	6	4	3	2	1	1	1
Precipitation	1	1	1	2	1	1	2	2
Vapor pressure deficit	1	1	1	1	1	1	2	3

Models transferred to predict yield in unknown regions or years

Models lost accuracy when transferred to unknown regions, particularly when they were trained with data from only a few other regions, but maintained accuracy when transferred to unknown years. For instance, when predicting region 4, the model trained with data exclusively from region 4 achieved a MAE of 0.5 t ha^{-1} at anthesis (Fig. 3.5, orange dashed line). In contrast, models trained with data from a single other region showed consistently higher errors, with MAE reaching up to 1.65 t ha^{-1} (Fig. 3.5, gray points), representing a 1.15 t ha^{-1} error increase attributable to spatial transfer (1.65 t ha^{-1} vs. 0.5 t ha^{-1}). However, when predicting the same region with a model trained using data from all the remaining six regions, the MAE was 0.72 t ha^{-1}, resulting in a considerably smaller error increase of 0.22 t ha^{-1} (0.72 t ha^{-1} vs. 0.5 t ha^{-1}). This pattern of higher errors when transferring models trained with data from only a few regions was observed across all regions, with some showing more pronounced increases (e.g., region 1 or 7) than others (e.g., region 3). In contrast, transferring models between years resulted in only slight error increases. For the worst case, predicting yield for 2022 with a model trained on other years resulted in a MAE of 0.92 t ha⁻¹, compared to 0.54 t ha⁻¹ when using data from 2022 itself, representing an error increase of 0.38 t ha⁻¹ due to temporal transfer (0.92 t ha^{-1} vs 0.54 t ha^{-1}). In summary, transferring models to predict unknown regions led to MAE increases that were three times higher in the worst case than those observed when transferring to unknown years (1.15 t ha^{-1} vs. 0.38 t ha^{-1}).

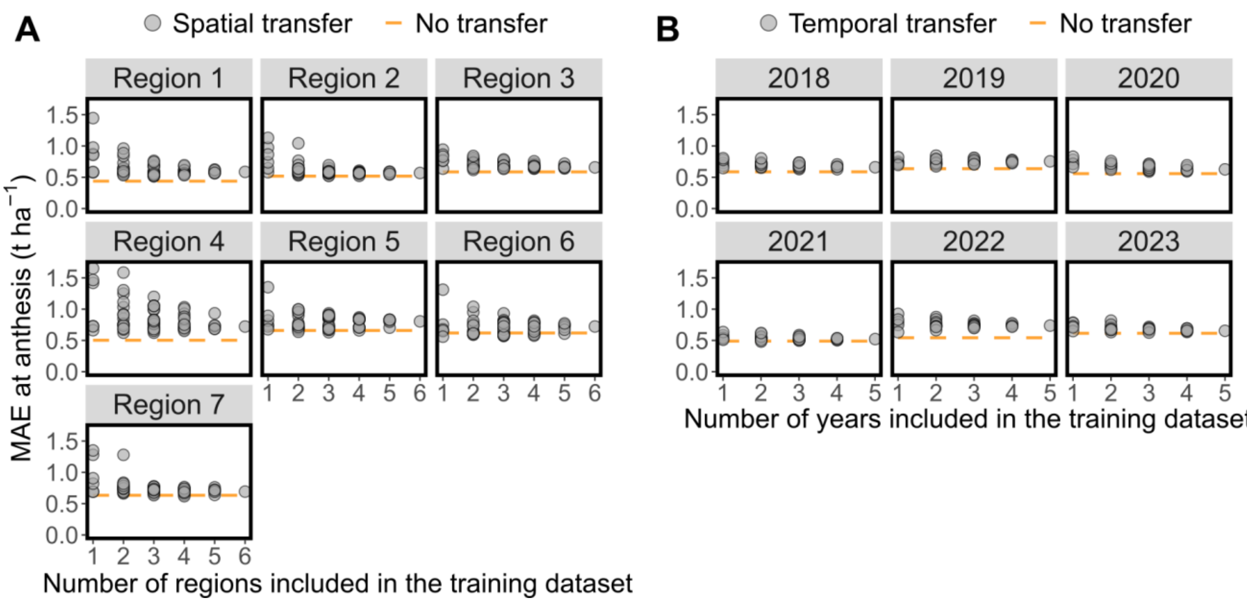

Figure 3.5: *Model accuracy when transferred to predict unknown regions or years. A) Median absolute error (MAE, t ha^{-1}) at anthesis as a function of the number of regions included in the training dataset. Each panel corresponds to a predicted region outside the training dataset. Gray points represent the MAE for specific combinations of regions used in model training. The orange dashed lines indicate the MAE for models trained and evaluated within the same region (see methodology for objective 1 and Fig. 3.3). For the location of each region, see Fig. 3.1A. B) Same as A), but for predicted years based on models trained with data from other years.*

As the largest error increases due to transfer were observed when training models with data from a single region or year, we analyzed each case in greater detail. On one hand, we identified regions and years with good general transferability. For example, models trained on region 1, the one with lowest yields, maintained acceptable accuracy when transferred to other regions (Fig. 3.6A). Similarly, models trained on data from 2021 achieved adequate accuracy when transferred to other years, except in the case of 2022 (Fig. 3.6B). On the other hand, we identified a phenomenon of "hysteresis", in which a model trained on a specific region or year retained accuracy when transferred to a second region or year, but lost accuracy when transferred from the second to the first. An example of this phenomenon is observed between regions 1 and 7: a model trained on region 1 successfully predicted region 7 with a MAE of 0.69 t ha^{-1}, only slightly higher than the MAE within region 7 (0.63 t ha^{-1}). However, when the model was trained on data from region 7 and transferred to region 1, the MAE increased to 1.44 t ha^{-1}, considerably higher than the MAE within region 1 (0.44 t ha^{-1}). In general, models were transferable when regions or years had a similar relationship between yield and fAPARgreen at anthesis, regardless of the distribution of yield or fAPARgreen data (examples in Fig. S6).

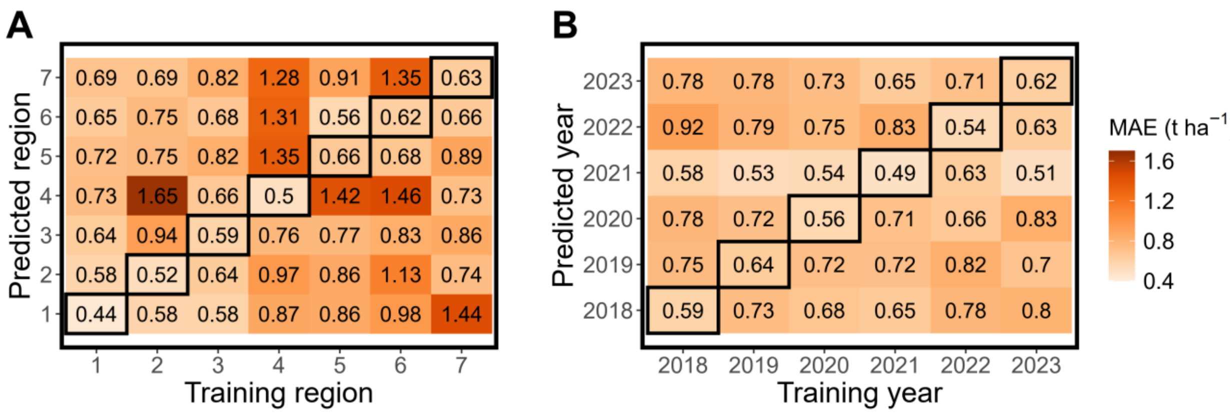

Figure 3.6: *Model accuracy when predicting yield with models trained on a single region (A) or year (B). A) Median absolute error (MAE, t ha^{-1}) at anthesis for predicted regions based on models trained using a single region. Values in the diagonal within black squares indicate the MAE at anthesis for models trained and evaluated within the same region (see methodology for objective 1 and Fig. 3.3.). For the location of each region, see Fig. 3.1A. B) Same as A), but for predicted years based on models trained using a single year.*

Models to predict yield in future years from the cumulative training from previous years.

Our models predicted yield at anthesis with varying accuracy when applied to future years, simulating operational conditions (Fig. 3.7). Between 2019 and 2021, accuracy progressively increased, coinciding with an increasing number of years available for model training, from only one year in 2019 (2018) to three years in 2021 (2018–2020). For instance, the MAE at anthesis was 0.75 t ha^{-1} in 2019, 0.67 t ha^{-1} in 2020, and 0.53 t ha^{-1} in 2021. Similarly, the MAPE at anthesis was 18% in 2019 and 2020, decreasing to 12% in 2021. However, this trend was interrupted in 2022, when the model recorded the lowest accuracy, with a MAE of 0.76 t ha^{-1} and MAPE of 24%. This year was highly atypical for wheat in Argentina due to repeated frosts during the crop's critical period near anthesis. Finally, in 2023, accuracy returned to normal values, with a MAE of 0.65 t ha^{-1} and MAPE of 16%.

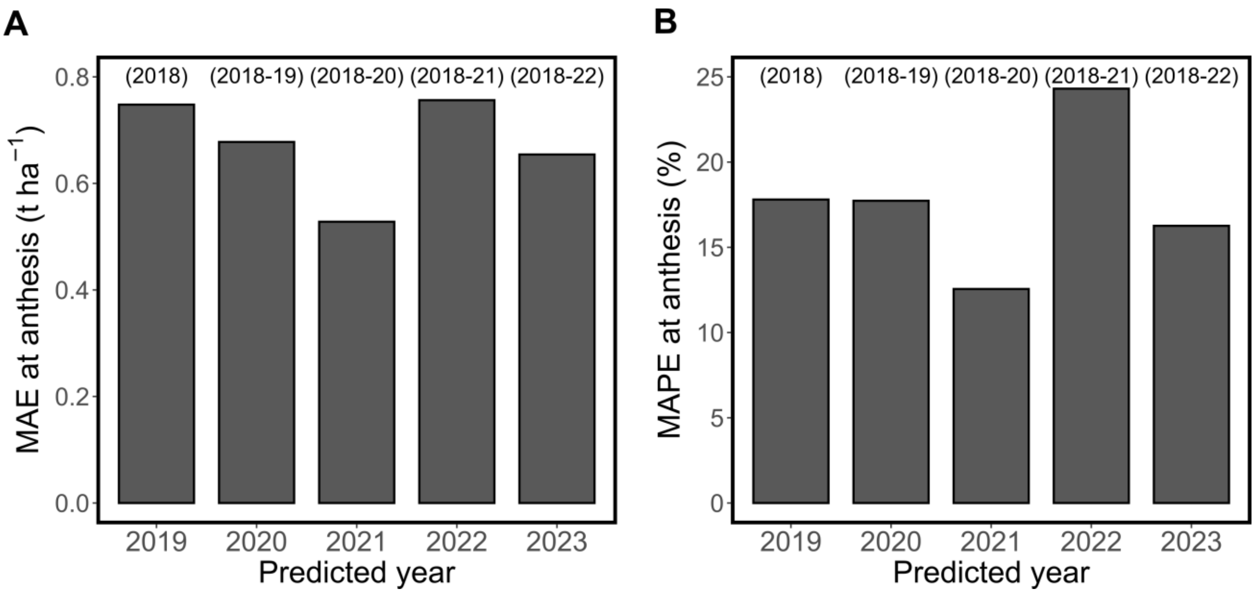

Figure 3.7: *Accuracy metrics for the simulation of operational conditions. A) Median absolute error at anthesis (MAE, t ha^{-1}) for each predicted year (2019 to 2023). The model for each year was trained using data from previous years, as indicated in parentheses above the columns. B) Same as A), but showing the median absolute percentage error at anthesis (MAPE, %).*

Discussion

This study provides new insights into the spatio-temporal transferability of crop yield prediction models using remote sensing. Leveraging a large field-scale database combined with satellite and weather data, we developed wheat yield models primarily driven by absorbed radiation (fAPARgreen). These models demonstrated good performance in predicting yield for unknown years but, when transferred to unknown regions, they lost accuracy. This reduction in accuracy is likely attributable from greater variability in the efficiency of converting absorbed radiation into grain yield (i.e., grain yield per unit of fAPARgreen) across different regions. Factors such as climatic conditions (Barlow et al., 2015; Rodriguez & Sadras, 2007), agronomic practices

(Caviglia & Sadras, 2001; Estrada-Campuzano et al., 2012), and genotypic differences (Miralles & Slafer, 1997; Reynolds et al., 2012) likely influence this efficiency in wheat. Supporting this rationale, training models with datasets that progressively included more regions mitigated this issue by capturing a more representative "average" conversion efficiency, reducing spatial transfer errors. Some studies (e.g., Gaso et al., 2025; Khaki et al., 2020) downplay the challenges of spatial transfer probably because they evaluated predictions within the geographic domain of the training data, unlike our approach of testing on entirely new regions. Other research (e.g., Ma et al., 2021; Priyatirkto et al., 2023) supports our finding that transferring models across regions is more difficult than transferring them across years.

Our findings about model transferability provide guidance for developing yield prediction systems. From a model developer's perspective, when resources for data collection—such as polygons, yield, or management data—are limited, our results suggest that prioritizing data from multiple regions over multiple years better captures variability for model training. On the other hand, from a user's perspective, models trained on data from a few years and applied within the same training region are likely to perform better than those trained on multiple years but transferred from other regions. Consequently, one primary goal of these systems—forecasting yield in future years—can be achieved once sufficient data from the target region has been collected. These models not only enable forecasting yield, but also mapping historical yields to analyze yield-determining factors and gaps relative to potential yields (Deines et al., 2023; Lobell, 2013).

For regions with limited or no data, transferring models from other regions remains a significant challenge. A potential solution involves developing agronomically interpretable models. By explicitly accounting for the factors driving yield predictions, these models enable developers to better anticipate performance in regions outside the training set (Loyola-Gonzalez, 2019; Rudin, 2019). Incorporating process-based knowledge helps generate hypotheses about the agronomic contexts where a model is likely to perform well or face limitations, as interpretability provides insights into how predictions are constructed. Furthermore, our findings underscore the need for rigorous transferability evaluations that reflect realistic error rates. These evaluations should mimic the conditions under which the models are intended to operate—such as forecasting in future years or predicting in previously unobserved regions—, ensuring accurate communication of real-world application errors (Fillippi et al., 2025; Morales & Villalobos, 2023). This approach helps avoid reliance on artificially low error rates often produced by conventional random dataset splitting, where training and evaluation datasets are not fully spatio-temporally independent (Roberts et al., 2017).

Our models predicted wheat yield well before harvest at the field scale, combining accuracy, simplicity, and interpretability. First, unlike most studies that report calibration errors or evaluate models using random data partitions or by predicting past years with future data, we assessed accuracy by predicting future years (Paudel et al., 2023; Schauberger et al., 2020). Second, our models provided field-scale predictions with broad spatial and temporal coverage, while most prior studies operate at the regional scale or rely on geographically concentrated datasets spanning limited years, with few exceptions (Brinkhoff et al., 2024; Deines et al., 2021; Ma et al., 2024). In fact, all previous work in our study region used county-scale (not field-scale) data (Lopresti et al., 2015; Zachow et al., 2024). Third, our models achieved acceptable errors 40–50 days before anthesis, with accuracy improving as anthesis approached. Early predictions can guide management decisions, such as late fertilization or fungicide applications, while predictions at anthesis support commercial and harvest decisions (Fieuzal & Baup, 2017; Shuai & Basso, 2022). Fourth, our models offer a simpler and more practical alternative to those needing extensive data inputs, only needing sowing date and field location as user inputs. Finally, their agronomic interpretability, based on absorbed radiation, makes them valuable for both yield prediction and post-season analysis.

Wheat yield predictions were primarily driven by fAPARgreen from satellite data, with accuracy improving as anthesis approached. During this stage, the crop determines its main numerical yield component, grain number, which is closely linked to the crop growth rate, explaining why fAPARgreen serves as a reliable predictor (Carrera et al., 2024). Before anthesis, accuracy depended on the correlation between early-stage fAPARgreen and its value at anthesis, along with deviations of final yield from the historical average. Early predictions largely reflected the historical average, minimally adjusted by fAPARgreen. After anthesis, during grain filling, accuracy remained stable since wheat yield variation is more proportionally defined by grain number rather than grain weight (Borrás et al., 2004; Slafer et al., 2014). This contrasts with crops like maize and soybean, in which post-flowering satellite data improved predictions (Lobell et al., 2015; Schwalbert et al., 2020). These results reinforce the importance of incorporating crop phenological estimates for accurate yield predictions using remote sensing (Bolton & Friedl., 2013; Nieto et al., 2022). Additionally, fAPARgreen proved to be far more relevant for yield prediction than weather variables, likely because it effectively captures weather effects through its relationship with crop growth. Furthermore, simpler algorithms like Elastic-Net, which leverage the linear relationship between yield and fAPARgreen, performed as well as nonlinear machine learning algorithms such as Random Forest or XGBoost (Appendix B, Fig. S7).

Future research could explore the following priorities: 1) extend transferability assessments to other major crops beyond wheat; 2) address spatial transferability across more contrasting global wheat producing regions; 3) expand temporal analyses beyond six years to better capture long-term variability driven by climate change, new genotypes, and evolving management practices; and 4) improve accessibility to field-scale weather data more related to conversion efficiency from absorbed radiation into yield, such as freeze or heat damage, to reduce satellite-based model error.

Conclusion

Yield prediction models were primarily driven by satellite-derived absorbed radiation (fAPARgreen), showed increasing accuracy as the crop cycle progressed toward anthesis. While the models retained accuracy when transferred to predict yield for unknown years, they lost it when applied to unknown regions. This result is likely attributable to greater spatial variability in the efficiency of converting absorbed radiation into yield. Notably, the models successfully predicted future-year yields at the field scale when trained on previous years, simulating operational conditions and achieving errors of 0.67 t ha^{-1} (17%) at anthesis, 50 days before harvest. This achievement is significant as the models required only field location and sowing date as user inputs. These findings contribute to advancing the development of more universal yield prediction models based on remote sensing.

4. Final remarks

In this thesis, we have revisited frameworks for crop yield prediction using the most widely used data source: spectral data from remote sensors. Among the available methods for yield prediction—including regional surveys, weather variables, and crop simulation models— remote sensing stands out for its ability to acquire detailed spatio-temporal crop information in a fast, accurate and non-destructive manner. This spectral information can be integrated into predictive models through various frameworks, each with different data requirements, accuracy, interpretability, and transferability to other agronomic scenarios.

Given this diversity, in the second chapter we synthesized the main frameworks for crop yield prediction based on spectral data. We identified three major groups of models, each with specific subgroups: (1) models based directly on spectral data, further classified into simple statistical models and machine learning models; (2) models based on derived biophysical variables, including empirical approaches and models based on resource capture and use efficiencies; and (3) models based on the integration of spectral data with crop simulation models, ranging from data assimilation to metamodel calibration and hybrid model development. For each framework, we described how spectral data are incorporated into predictive models and analyzed three key aspects: data requirements for calibration, the trade-off between interpretability and predictive power, and spatio-temporal transferability. In this synthesis, we did not establish a single superior framework, but highlighted that the optimal choice depends on the objective, data availability and expertise of the researcher.

Building on this knowledge, in the third chapter we developed yield prediction models for wheat at the field-scale, requiring only sowing date and location as user inputs. These models relied primarily on satellite-derived absorbed radiation, supplemented by weather data. To assess the spatio-temporal transferability of the models, our main focus, we implemented three predictive approaches. In the first approach, we trained and evaluated models within the same region or year and obtained the lowest prediction error at anthesis, approximately 50 days before harvest (0.61 t ha^{-1}; 16%). In the second approach, we evaluated models in unknown regions or years. The models maintained accuracy when predicting in unknown years but lost accuracy when applied to unknown regions, with spatial transfer errors up to three times larger than temporal transfer errors. In the third approach, we simulated operational conditions by predicting future years using models trained exclusively on past years, yielding a median error at anthesis of 0.67 t ha^{-1} (17%). Our results demonstrated the robustness of satellite-based models for yield prediction in unknown years while highlighting the challenges associated with spatial transferability. As these models were based on absorbed radiation, some of the error was likely due to variations in the efficiency of converting absorbed radiation into yield, which appeared to vary more between regions than between years. In summary, these field-scale models combined accuracy, pre-harvest prediction and simplicity, making them a valuable tool for a range of stakeholders.

Future research could expand the findings of this thesis. In particular, we identify three promising research directions: (1) comparing multiple yield prediction frameworks within the same study and dataset. These include simple regressions between yield and VIs, machine learning and deep learning algorithms, and models that integrate spectral data with simulation models. A key aspect is evaluating how interpretability, accuracy, and transferability change depending on the amount and level of data processing (e.g., raw spectral bands, VIs, or derived biophysical variables); (2) assessing model transferability in more challenging scenarios, such as their application across different production systems or crops, to develop more universal models; and (3) incorporating complementary remote sensing data sources, such as solar-induced chlorophyll fluorescence (SIF), to better capture the effect of crop conversion efficiency. Unlike traditional VIs, which estimate the radiation absorbed by the crop, SIF captures variations in the efficiency of the crop to convert absorbed radiation into biomass. Although some studies have explored this index (e.g., Cai et al. 2019), its application in yield prediction remains limited due to the relatively low spatio-temporal resolution of available satellite-based measurements. Advancing these research directions would contribute to the development of more accurate, interpretable, and universal models.

5. References

Andrade, F. H., Uhart, S. A., Arguissain, G. G., & Ruiz, R. A. (1992). Radiation use efficiency of maize grown in a cool area. Field Crops Research, 28(4), 345-354.

Ardagh, J. C. (1889). Nilometers. In Proceedings of the Royal Geographical Society and Monthly Record of Geography (Vol. 11, No. 1, pp. 28-38). Royal Geographical Society (with the Institute of British Geographers), Wiley.

Azzari, G., Jain, M., & Lobell, D. B. (2017). Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries. Remote Sensing of Environment, 202, 129-141.

Barlow, K. M., Christy, B. P., O'Leary, G. J., Riffkin, P. A., & Nuttall, J. G. (2015). Simulating the impact of extreme heat and frost events on wheat crop production: A review. Field crops research, 171, 109-119.

Basso, B., & Liu, L. (2019). Seasonal crop yield forecast: Methods, applications, and accuracies. Advances in agronomy, 154, 201-255.

Bastiaanssen, W. G., & Ali, S. (2003). A new crop yield forecasting model based on satellite measurements applied across the Indus Basin, Pakistan. Agriculture, ecosystems & environment, 94(3), 321-340.

Battude, M., Al Bitar, A., Morin, D., Cros, J., Huc, M., Sicre, C. M., ... & 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.

Becker-Reshef, I., Vermote, E., Lindeman, M., & Justice, C. (2010). A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. Remote sensing of environment, 114(6), 1312-1323.

Benami, E., Jin, Z., Carter, M. R., Ghosh, A., Hijmans, R. J., Hobbs, A., ... & Lobell, D. B. (2021). Uniting remote sensing, crop modelling and economics for agricultural risk management. Nature Reviews Earth & Environment, 2(2), 140-159.

Bolton, D. K., & Friedl, M. A. (2013). Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agricultural and forest meteorology, 173, 74-84.

Borrás, L., Slafer, G. A., & Otegui, M. E. (2004). Seed dry weight response to source–sink manipulations in wheat, maize and soybean: a quantitative reappraisal. Field Crops Research, 86(2-3), 131-146.

Bouman, B. A. M. (1995). Crop modelling and remote sensing for yield prediction. Netherlands journal of agricultural science, 43, 143-143.

Brinkhoff, J., Clarke, A., Dunn, B. W., & Groat, M. (2024). Analysis and forecasting of Australian rice yield using phenology-based aggregation of satellite and weather data. Agricultural and Forest Meteorology, 353, 110055.

Burke, M., Driscoll, A., Lobell, D. B., Ermon, S. (2020). Using satellite imagery to understand and promote sustainable development. 1219(March). https://doi.org/10.1126/science.abe8628

Cai, Y., Guan, K., Lobell, D., Potgieter, A. B., Wang, S., Peng, J., ... & Peng, B. (2019). Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agricultural and forest meteorology, 274, 144-159.

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.

Calviño, P., & Monzon, J. P. (2009). Farming systems of Argentina: yield constraints and risk management. Crop physiology: Applications for genetic improvement and agronomy, 51, 70.

Campoy, J., Campos, I., Villodre, J., Bodas, V., Osann, A., & Calera, A. (2023). Remote Sensing-based crop yield model at field and within-field scales in wheat and barley crops. European Journal of Agronomy, 143, 126720.

Cao, J., Zhang, Z., Tao, F., Zhang, L., Luo, Y., Zhang, J., ... & Xie, J. (2021). Integrating multi-source data for rice yield prediction across China using machine learning and deep learning approaches. Agricultural and Forest Meteorology, 297, 108275.

Carrera, C. S., Savin, R., & Slafer, G. A. (2024). Critical period for yield determination across grain crops. Trends in Plant Science, 29(3), 329-342.

Caviglia, O. P., & Sadras, V. O. (2001). Effect of nitrogen supply on crop conductance, water- and radiation-use efficiency of wheat. Field Crops Research, 69(3), 259-266.

Chen, Y., & Tao, F. (2020). Improving the practicability of remote sensing data-assimilation-based crop yield estimations over a large area using a spatial assimilation algorithm and ensemble assimilation strategies. Agricultural and Forest Meteorology, 291, 108082.

Chen, Y., Donohue, R. J., McVicar, T. R., Waldner, F., Mata, G., Ota, N., ... & Lawes, R. A. (2020). Nationwide crop yield estimation based on photosynthesis and meteorological stress indices. Agricultural and Forest Meteorology, 284, 107872.

Cheng, Z., Gu, X., Zhou, Z., Zhang, Y., Yin, H., Li, W., ... & Du, Y. (2024). Enhancing in-season yield forecast accuracy for film-mulched wheat: A hybrid approach coupling crop model and UAV remote-sensing data by ensemble learning technique. European Journal of Agronomy, 156, 127174.

Chipanshi, A. C., Ripley, E. A., Lawford, R. G. (1997). Early prediction of spring wheat yields in Saskatchewan from current and historical weather data using the CERES-Wheat model. Agricultural and Forest Meteorology, 84(3-4), 223-232.

Chipanshi, A., Zhang, Y., Kouadio, L., Newlands, N., Davidson, A., Hill, H., ... & 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.

Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Computers and electronics in agriculture, 151, 61-69.

Clark, R., Dahlhaus, P., Robinson, N., Larkins, J. A., & Morse-McNabb, E. (2023). Matching the model to the available data to predict wheat, barley, or canola yield: A review of recently published models and data. Agricultural Systems, 211, 103749.

Clevers, J. G. P. W. (1997). A simplified approach for yield prediction of sugar beet based on optical remote sensing data. Remote sensing of Environment, 61(2), 221-228.

Clevers, J., Vonder, O., Jongschaap, R., Desprats, J. F., King, C., Prévot, L., & Bruguier, N. (2002). Using SPOT data for calibrating a wheat growth model under mediterranean conditions. Agronomie, 22(6), 687-694.

Colombo, R., Bellingeri, D., Fasolini, D., & Marino, C. M. (2003). Retrieval of leaf area index in different vegetation types using high resolution satellite data. Remote sensing of environment, 86(1), 120-131.

Colwell, J. E., Rice, D. P., & Nalepka, R. F. (1977). Wheat yield forecasts using Landsat data. In Proc. of the 11th Intern. Symp. on Remote Sensing of Environment, Vol. 2.

Curnel, Y., de Wit, A. J., Duveiller, G., & Defourny, P. (2011). Potential performances of remotely sensed LAI assimilation in WOFOST model based on an OSS Experiment. Agricultural and forest meteorology, 151(12), 1843-1855.

Darvishzadeh, R., Skidmore, A., Schlerf, M., & Atzberger, C. (2008). Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. Remote sensing of environment, 112(5), 2592-2604.

Deines, J. M., Archontoulis, S. V., Huber, I., & Lobell, D. B. (2024). Observational evidence for groundwater influence on crop yields in the United States. Proceedings of the National Academy of Sciences, 121(36), e2400085121.

Deines, J. M., Guan, K., Lopez, B., Zhou, Q., White, C. S., Wang, S., & Lobell, D. B. (2023). Recent cover crop adoption is associated with small maize and soybean yield losses in the United States. Global change biology, 29(3), 794-807.

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.

Dhakar, R., Sehgal, V. K., Chakraborty, D., Sahoo, R. N., Mukherjee, J., Ines, A. V., ... & Roy, S. B. (2022). Field scale spatial wheat yield forecasting system under limited field data availability by integrating crop simulation model with weather forecast and satellite remote sensing. Agricultural Systems, 195, 103299.

Duan, B., Fang, S., Gong, Y., Peng, Y., Wu, X., & Zhu, R. (2021). Remote estimation of grain yield based on UAV data in different rice cultivars under contrasting climatic zone. Field Crops Research, 267, 108148.

Duchon, C. E. (1986). Corn yield prediction using climatology. Journal of Applied Meteorology and Climatology, 25(5), 581-590.

Epiphanio, J. N., & Huete, A. R. (1995). Dependence of NDVI and SAVI on sun/sensor geometry and its effect on fAPAR relationships in Alfalfa. Remote Sensing of Environment, 51(3), 351-360.

Estrada-Campuzano, G., Slafer, G. A., & Miralles, D. J. (2012). Differences in yield, biomass and their components between triticale and wheat grown under contrasting water and nitrogen environments. Field Crops Research, 128, 167-179.

Feng, P., Wang, B., Li Liu, D., Waters, C., Xiao, D., Shi, L., & Yu, Q. (2020). Dynamic wheat yield forecasts are improved by a hybrid approach using a biophysical model and machine learning technique. Agricultural and Forest Meteorology, 285, 107922.

Fieuzal, R., & Baup, F. (2017). Forecast of wheat yield throughout the agricultural season using optical and radar satellite images. International Journal of Applied Earth Observation and Geoinformation, 59, 147-156.

Filippi, P., Han, S. Y., & Bishop, T. F. (2025). On crop yield modelling, predicting, and forecasting and addressing the common issues in published studies. Precision Agriculture, 26(1), 1-19.

Filippi, P., Jones, E. J., Wimalathunge, N. S., Somarathna, P. D., Pozza, L. E., Ugbaje, S. U., ... & Bishop, T. F. (2019). An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. Precision Agriculture, 20, 1015-1029.

Foulkes, M. J., Slafer, G. A., Davies, W. J., Berry, P. M., Sylvester-Bradley, R., Martre, P., Calderini, D. F., Griffiths, S. & Reynolds, M. P. (2011). Raising yield potential of wheat. III. Optimizing partitioning to grain while maintaining lodging resistance. Journal of experimental botany, 62(2), 469-486.

Friedman, J., Hastie, T., Tibshirani, R., Narasimhan, B., Tay, K., Simon, N., & Qian, J. (2021). Package ‘glmnet’. CRAN R Repositary, 595.

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(May), 258–272. https://doi.org/10.1016/j.agsy.2018.05.010

Gallo, K. P., Daughtry, C. S. T., & Bauer, M. E. (1985). Spectral estimation of absorbed photosynthetically active radiation in corn canopies. Remote Sensing of Environment, 17(3), 221-232.

Gao, F., Anderson, M. C., Zhang, X., Yang, Z., Alfieri, J. G., Kustas, W. P., ... & Prueger, J. H. (2017). Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. Remote Sensing of Environment, 188, 9-25.

Garbulsky, M. F., Peñuelas, J., Gamon, J., Inoue, Y., & Filella, I. (2011). The photochemical reflectance index (PRI) and the remote sensing of leaf, canopy and ecosystem radiation use efficiencies: A review and meta-analysis. Remote sensing of environment, 115(2), 281-297.

Gaso, D. V., de Wit, A., Berger, A. G., & Kooistra, L. (2021). Predicting within-field soybean yield variability by coupling Sentinel-2 leaf area index with a crop growth model. Agricultural and forest meteorology, 308, 108553.

Gaso, D. V., de Wit, A., de Bruin, S., Puntel, L. A., Berger, A. G., & Kooistra, L. (2023). Efficiency of assimilating leaf area index into a soybean model to assess within-field yield variability. European journal of agronomy, 143, 126718.

Gaso, D. V., La Rosa, L. E. C., Puntel, L. A., Edreira, J. I. R., de Wit, A., & Kooistra, L. (2025). Predicting spatial and temporal variability in soybean yield using deep learning and open source data. European Journal of Agronomy, 164, 127498.

Gaso, D. V., Paudel, D., de Wit, A., Puntel, L. A., Mullissa, A., & Kooistra, L. (2024). Beyond assimilation of leaf area index: Leveraging additional spectral information using machine learning for site-specific soybean yield prediction. Agricultural and Forest Meteorology, 351, 110022.

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

Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. Remote sensing of Environment, 80(1), 76-87.

Gitelson, A. A., Peng, Y., & Huemmrich, K. F. (2014). Relationship between fraction of radiation absorbed by photosynthesizing maize and soybean canopies and NDVI from remotely sensed data taken at close range and from MODIS 250 m resolution data. Remote Sensing of Environment, 147, 108-120.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote sensing of Environment, 202, 18-27.

Groten, S.M.E., (1993). NDVI—crop monitoring and early yield assessment of Burkina-Faso. Int. J. Remote Sens. 14, 1495–1515.

Hammond, A. L. (1975). Crop forecasting from space: toward a global food watch. Science, 188(4187), 434-436.

Han, C., Zhang, B., Chen, H., Wei, Z., & Liu, Y. (2019). Spatially distributed crop model based on remote sensing. Agricultural Water Management, 218, 165-173.

Han, J., Zhang, Z., Cao, J., Luo, Y., Zhang, L., Li, Z., & Zhang, J. (2020). Prediction of winter wheat yield based on multi-source data and machine learning in China. Remote Sensing, 12(2), 236.

Hassan, M. A., Yang, M., Rasheed, A., Yang, G., Reynolds, M., Xia, X., ... & He, Z. (2019). A rapid monitoring of NDVI across the wheat growth cycle for grain yield prediction using a multi-spectral UAV platform. Plant science, 282, 95-103.

Heilman, J. L., Kanemasu, E. T., Bagley, J. O., & Rasmussen, V. P. (1977). Evaluating soil moisture and yield of winter wheat in the Great Plains using Landsat data. Remote sensing of Environment, 6(4), 315-326.

Hilker, T., Coops, N. C., Wulder, M. A., Black, T. A., & Guy, R. D. (2008). The use of remote sensing in light use efficiency based models of gross primary production: A review of current status and future requirements. Science of the Total Environment, 404(2-3), 411-423.

Hodges, T., Botner, D., Sakamoto, C., & Haug, J. H. (1987). Using the CERES-Maize model to estimate production for the US Cornbelt. Agricultural and Forest Meteorology, 40(4), 293-303.

Homolová, L., Malenovský, Z., Clevers, J. G., García-Santos, G., & Schaepman, M. E. (2013). Review of optical-based remote sensing for plant trait mapping. Ecological Complexity, 15, 1-16.

Hu, T., Zhang, X., Bohrer, G., Liu, Y., Zhou, Y., Martin, J., ... & Zhao, K. (2023). Crop yield prediction via explainable AI and interpretable machine learning: Dangers of black box models for evaluating climate change impacts on crop yield. Agricultural and Forest Meteorology, 336, 109458.

Huang, J., Gómez-Dans, J. L., Huang, H., Ma, H., Wu, Q., Lewis, P. E., ... & Xie, X. (2019). Assimilation of remote sensing into crop growth models: Current status and perspectives. Agricultural and forest meteorology, 276, 107609.

Huang, J., Ma, H., Sedano, F., Lewis, P., Liang, S., Wu, Q., ... & Zhu, D. (2019). Evaluation of regional estimates of winter wheat yield by assimilating three remotely sensed reflectance datasets into the coupled WOFOST–PROSAIL model. European journal of agronomy, 102, 1-13.

Huang, J., Sedano, F., Huang, Y., Ma, H., Li, X., Liang, S., ... & Wu, W. (2016). Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to improve regional winter wheat yield estimation. Agricultural and Forest Meteorology, 216, 188-202.

Jin, Z., Azzari, G., & Lobell, D. B. (2017). Improving the accuracy of satellite-based high-resolution yield estimation: A test of multiple scalable approaches. Agricultural and forest meteorology, 247, 207-220.

Johnson, D. M. (2014). An assessment of pre-and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. Remote Sensing of Environment, 141, 116-128.

Kamir, E., Waldner, F., & Hochman, Z. (2020). Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. ISPRS Journal of Photogrammetry and Remote Sensing, 160, 124-135.

Kang, Y., & Özdoğan, M. (2019). Field-level crop yield mapping with Landsat using a hierarchical data assimilation approach. Remote sensing of environment, 228, 144-163.

Kemanian, A.R., Stockle, C.O., Huggins, D.R., Viega, L.M., (2007). A simple method to estimate harvest index in grain crops. Field Crops Res 103, 208–216. https://doi.org/10.1016/j.fcr.2007.06.007.

Khaki, S., Wang, L., & Archontoulis, S. V. (2020). A CNN-RNN framework for crop yield prediction. Frontiers in Plant Science, 10, 1750.

Kolotii, A., Kussul, N., Shelestov, A., Skakun, S., Yailymov, B., Basarab, R., Lavreniuk, M., Oliinyk, T., Ostapenko, V., 2015. Comparison of biophysical and satellite predictors for wheat yield forecasting in Ukraine. In: Schreier, G., Skrovseth, P.E., Staudenrausch, H. (Eds.), 36th International Symposium on Remote Sensing of Environment. vol. 47., pp. 39–44.

Lai, Y., Pringle, M. J., Kopittke, P. M., Menzies, N. W., Orton, T. G., & Dang, Y. P. (2018). An empirical model for prediction of wheat yield, using time-integrated Landsat NDVI. International journal of applied earth observation and geoinformation, 72, 99-108.

Lal, R. (2016). Feeding 11 billion on 0.5 billion hectare of area under cereal crops. Food and Energy Security, 5(4), 239-251.

Li, J., Li, G., Wang, L., Li, D., Li, H., Gao, C., ... & Wang, E. (2023). Predicting maize yield in Northeast China by a hybrid approach combining biophysical modelling and machine learning. Field Crops Research, 302, 109102.

Li, Y., Zeng, H., Zhang, M., Wu, B., Zhao, Y., Yao, X., ... & Wu, F. (2023). A county-level soybean yield prediction framework coupled with XGBoost and multidimensional feature engineering. International Journal of Applied Earth Observation and Geoinformation, 118, 103269.

Lobell, D. B. (2013). The use of satellite data for crop yield gap analysis. Field crops research, 143, 56-64.

Lobell, D. B., Asner, G. P., Ortiz-Monasterio, J. I., & Benning, T. L. (2003). Remote sensing of regional crop production in the Yaqui Valley, Mexico: estimates and uncertainties. Agriculture, Ecosystems & Environment, 94(2), 205-220.

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.

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

Lopresti, M. F., Di Bella, C. M., & Degioanni, A. J. (2015). Relationship between MODIS-NDVI data and wheat yield: A case study in Northern Buenos Aires province, Argentina. Information Processing in Agriculture, 2(2), 73-84.

Loyola-Gonzalez, O. (2019). Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view. IEEE access, 7, 154096-154113.

Ma, Y., Liang, S. Z., Myers, D. B., Swatantran, A., & Lobell, D. B. (2024). Subfield-level crop yield mapping without ground truth data: A scale transfer framework. Remote Sensing of Environment, 315, 114427.

Ma, Y., Yang, Z., Huang, Q., & Zhang, Z. (2023). Improving the transferability of deep learning models for crop yield prediction: A partial domain adaptation approach. Remote Sensing, 15(18), 4562.

Ma, Y., Zhang, Z., Kang, Y., & Özdoğan, M. (2021). Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. Remote Sensing of Environment, 259, 112408.

Ma, Y., Zhang, Z., Yang, H. L., & Yang, Z. (2021). An adaptive adversarial domain adaptation approach for corn yield prediction. Computers and Electronics in Agriculture, 187, 106314.

MacDonald, R. B., & Hall, F. G. (1980). Global crop forecasting. Science, 208(4445), 670-679.

Maestrini, B., Mimić, G., van Oort, P. A., Jindo, K., Brdar, S., Athanasiadis, I. N., & van Evert, F. K. (2022). Mixing process-based and data-driven approaches in yield prediction. European Journal of Agronomy, 139, 126569.

Magney, T. S., Eitel, J. U., Huggins, D. R., & Vierling, L. A. (2016). Proximal NDVI derived phenology improves in-season predictions of wheat quantity and quality. Agricultural and Forest Meteorology, 217, 46-60.

Magrin, G. O., Travasso, M. I., & Rodríguez, G. R. (2005). Changes in climate and crop production during the 20th century in Argentina. Climatic change, 72, 229-249.

Manivasagam, V. S., & Rozenstein, O. (2020). Practices for upscaling crop simulation models from field scale to large regions. Computers and Electronics in Agriculture, 175, 105554.

Miralles, D. J., & Slafer, G. A. (1997). Radiation interception and radiation use efficiency of near-isogenic wheat lines with different height. Euphytica, 97, 201-208.

Miralles, D. J., Spinedi, M. V., Abeledo, L. G., & Abelleyra, D. (2007). Variability on photoperiod responses in Argentinean wheat cultivars differing in length of crop cycle. In Wheat Production in Stressed Environments: Proceedings of the 7th International Wheat Conference, 27 November–2 December 2005, Mar del Plata, Argentina (pp. 599-609). Springer Netherlands.

Mkhabela, M. S., Bullock, P., Raj, S., Wang, S., & Yang, Y. (2011). Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. Agricultural and Forest Meteorology, 151(3), 385-393.

Mohammed, G. H., Colombo, R., Middleton, E. M., Rascher, U., Van Der Tol, C., Nedbal, L., ... & Zarco-Tejada, P. J. (2019). Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress. Remote sensing of environment, 231, 111177.

Monteith, J.L., (1972). Solar radiation and productivity in tropical ecosystems. J. Appl. Ecol. 9, 747–766.

Morales, A., & Villalobos, F. J. (2023). Using machine learning for crop yield prediction in the past or the future. Frontiers in Plant Science, 14, 1128388.

Morel, J., Bégué, A., Todoroff, P., Martiné, J. F., Lebourgeois, V., & Petit, M. (2014). Coupling a sugarcane crop model with the remotely sensed time series of fIPAR to optimise the yield estimation. European journal of agronomy, 61, 60-68.

Mullissa, A., Vollrath, A., Odongo-Braun, C., Slagter, B., Balling, J., Gou, Y., ... & Reiche, J. (2021). Sentinel-1 sar backscatter analysis ready data preparation in google earth engine. Remote Sensing, 13(10), 1954.

Muruganantham, P., Wibowo, S., Grandhi, S., Samrat, N. H., & Islam, N. (2022). A systematic literature review on crop yield prediction with deep learning and remote sensing. Remote Sensing, 14(9), 1990.

Nandram, B., Berg, E., & Barboza, W. (2014). A hierarchical Bayesian model for forecasting state-level corn yield. Environmental and ecological statistics, 21, 507-530.

National Agricultural Statistics Service. Yield forecasting program of NASS. United States Department of Agriculture (2023).

Naylor, R. L., Falcon, W. P., Rochberg, D., & Wada, N. (2001). Using El Nino/Southern Oscillation climate data to predict rice production in Indonesia. Climatic Change, 50, 255-265.

Nevavuori, P., Narra, N., & Lipping, T. (2019). Crop yield prediction with deep convolutional neural networks. Computers and electronics in agriculture, 163, 104859.

Nguy-Robertson, A., Gitelson, A., Peng, Y., Viña, A., Arkebauer, T., & Rundquist, D. (2012). Green leaf area index estimation in maize and soybean: Combining vegetation indices to achieve maximal sensitivity. Agronomy journal, 104(5), 1336-1347.

Nicholls, N. (1986). Use of the Southern Oscillation to predict Australian sorghum yield. Agricultural and Forest Meteorology, 38(1-3), 9-15.

Nieto, L., Houborg, R., Zajdband, A., Jumpasut, A., Prasad, P. V., Olson, B. J., & Ciampitti, I. A. (2022). Impact of high-cadence earth observation in maize crop phenology classification. Remote Sensing, 14(3), 469.

Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R. L., & Mouazen, A. M. (2016). Wheat yield prediction using machine learning and advanced sensing techniques. Computers and electronics in agriculture, 121, 57-65.

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.

Paudel, D., De Wit, A., Boogaard, H., Marcos, D., Osinga, S., & Athanasiadis, I. N. (2023). Interpretability of deep learning models for crop yield forecasting. Computers and Electronics in Agriculture, 206, 107663.

Paveley, N. D., Lockley, K. D., Sylvester-Bradley, R., & Thomas, J. (1997). Determinants of fungicide spray decisions for wheat. Pesticide Science, 49(4), 379-388.

Pei, J., Tan, S., Zou, Y., Liao, C., He, Y., Wang, J., ... & Huang, J. (2025). The role of phenology in crop yield prediction: Comparison of ground-based phenology and remotely sensed phenology. Agricultural and Forest Meteorology, 361, 110340.

Pellegrini, P., Cossani, C. M., Bella, C. M. D., Piñeiro, G., Sadras, V. O., & Oesterheld, M. (2020). Simple regression models to estimate light interception in wheat crops with Sentinel-2 and a handheld sensor. Crop Science, 60(3), 1607-1616.

Peralta, N. R., Assefa, Y., Du, J., Barden, C. J., & Ciampitti, I. A. (2016). Mid-season high-resolution satellite imagery for forecasting site-specific corn yield. Remote Sensing, 8(10), 848.

Prince, S. D. (1991). A model of regional primary production for use with coarse resolution satellite data. International Journal of Remote Sensing, 12(6), 1313-1330.

Priyatikanto, R., Lu, Y., Dash, J., & Sheffield, J. (2023). Improving generalisability and transferability of machine-learning-based maize yield prediction model through domain adaptation. Agricultural and Forest Meteorology, 341, 109652.

QGIS Development Team, 2022. QGIS Geographic Information System. Open Source Geospatial Foundation. URL http://qgis.org

R Core Team (2023). _R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Rembold, F., Atzberger, C., Savin, I., & Rojas, O. (2013). Using low resolution satellite imagery for yield prediction and yield anomaly detection. Remote Sensing, 5(4), 1704-1733.

Ren, J., Chen, Z., Zhou, Q., & Tang, H. (2008). Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China. International Journal of Applied Earth Observation and Geoinformation, 10(4), 403-413.

Reynolds, M., Foulkes, J., Furbank, R., Griffiths, S., King, J., Murchie, E., Parry, M. & Slafer, G. (2012). Achieving yield gains in wheat. Plant, cell & environment, 35(10), 1799-1823.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., ... & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, 40(8), 913-929.

Rodriguez, D., & Sadras, V. O. (2007). The limit to wheat water-use efficiency in eastern Australia. I.* Gradients in the radiation environment and atmospheric demand. Australian Journal of Agricultural Research, 58(4), 287-302.

Rosenzweig, C., Jones, J. W., Hatfield, J. L., Ruane, A. C., Boote, K. J., Thorburn, P., ... & Winter, J. M. (2013). The agricultural model intercomparison and improvement project (AgMIP): protocols and pilot studies. Agricultural and forest meteorology, 170, 166-182.

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. NASA Spec. Publ, 351(1), 309.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature machine intelligence, 1(5), 206-215.

Sadras, V.O., Connor, D.J., 1991. Physiological basis of the response of harvest index to the fraction of water transpired after anthesis: a simple model to estimate harvest index for determinate species. Field Crops Res 26, 227–239. https://doi.org/10.1016/0378-4290(91)90001-C

Schauberger, B., Jägermeyr, J., & Gornott, C. (2020). A systematic review of local to regional yield forecasting approaches and frequently used data resources. European Journal of Agronomy, 120, 126153.

Schwalbert, R. A., Amado, T., Corassa, G., Pott, L. P., Prasad, P. V., & Ciampitti, I. A. (2020). Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. Agricultural and Forest Meteorology, 284, 107886.

Sellers, P. J. (1987). Canopy reflectance, photosynthesis, and transpiration, II. The role of biophysics in the linearity of their interdependence. Remote sensing of Environment, 21(2), 143-183.

Shanahan, J. F., Schepers, J. S., Francis, D. D., Varvel, G. E., Wilhelm, W. W., Tringe, J. M., ... & Major, D. J. (2001). Use of remote-sensing imagery to estimate corn grain yield. Agronomy journal, 93(3), 583-589.

Shewry, P. R., & Hey, S. J. (2015). The contribution of wheat to human diet and health. Food and energy security, 4(3), 178-202.

Shuai, G., & Basso, B. (2022). Subfield maize yield prediction improves when in-season crop water deficit is included in remote sensing imagery-based models. Remote Sensing of Environment, 272, 112938.

Sinclair, T. R., & Muchow, R. C. (1999). Radiation use efficiency. Advances in agronomy, 65, 215-265.

Slafer, G. A., Savin, R., & Sadras, V. O. (2014). Coarse and fine regulation of wheat yield components in response to genotype and environment. Field Crops Research, 157, 71-83.

Soler, C. M. T., Sentelhas, P. C., & Hoogenboom, G. (2007). Application of the CSM-CERES-Maize model for planting date evaluation and yield forecasting for maize grown off-season in a subtropical environment. European Journal of Agronomy, 27(2-4), 165-177.

Sun, J., Di, L., Sun, Z., Shen, Y., & Lai, Z. (2019). County-level soybean yield prediction using deep CNN-LSTM model. Sensors, 19(20), 4363.

Tao, F., Yokozawa, M., Zhang, Z., Xu, Y., & Hayashi, Y. (2005). Remote sensing of crop production in China by production efficiency models: models comparisons, estimates and uncertainties. Ecological modelling, 183(4), 385-396.

Thompson, L. M. (1969). Weather and Technology in the Production of Corn in the US Corn Belt 1. Agronomy Journal, 61(3), 453-456.

Trombetta, A., Iacobellis, V., Tarantino, E., & Gentile, F. (2016). Calibration of the AquaCrop model for winter wheat using MODIS LAI images. Agricultural Water Management, 164, 304-316.

Tucker, C. J., Holben, B. N., Elgin, J. J., & McMurtrey, J. I. (1981). Relationship of spectral data to grain yield variation. Photogrammetric Engineering and Remote Sensing, 46, 657-666.

Unkovich, M., Baldock, J., & Forbes, M. (2010). Variability in harvest index of grain crops and potential significance for carbon accounting: examples from Australian agriculture. Advances in agronomy, 105, 173-219.

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.

Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J. F., & Ceschia, E. (2017). Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. Remote sensing of environment, 199, 415-426.

Wang, J., Si, H., Gao, Z., & Shi, L. (2022). Winter wheat yield prediction using an LSTM model from MODIS LAI products. Agriculture, 12(10), 1707.

Wang, J., Wang, P., Tian, H., Tansey, K., Liu, J., & Quan, W. (2023). A deep learning framework combining CNN and GRU for improving wheat yield estimates using time series remotely sensed multi-variables. Computers and Electronics in Agriculture, 206, 107705.

Wang, L., Tian, Y., Yao, X., Zhu, Y., & Cao, W. (2014). Predicting grain yield and protein content in wheat by fusing multi-sensor and multi-temporal remote-sensing images. Field Crops Research, 164, 178-188.

Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. Remote sensing of Environment, 236, 111402.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., ... & Yutani, H. (2019). Welcome to the Tidyverse. Journal of open source software, 4(43), 1686.

Yang, Q., Shi, L., Han, J., Zha, Y., Yu, J., Wu, W., & Huang, K. (2023). Regulating the time of the crop model clock: A data assimilation framework for regions with high phenological heterogeneity. Field Crops Research, 293, 108847.

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

Yu, W., Li, D., Zheng, H., Yao, X., Zhu, Y., Cao, W., ... & Zhou, Y. (2024). HIDYM: a high-resolution gross primary productivity and dynamic harvest index based crop yield mapper. Remote Sensing of Environment, 311, 114301.

Yuan, W., Chen, Y., Xia, J., Dong, W., Magliulo, V., Moors, E., ... & Zhang, H. (2016). Estimating crop yield using a satellite-based light use efficiency model. Ecological Indicators, 60, 702-709.

Zachow, M., Kunstmann, H., Miralles, D. J., & Asseng, S. (2024). Multi-model ensembles for regional and national wheat yield forecasts in Argentina. Environmental Research Letters, 19(8), 084037.

Zare, H., Weber, T. K., Ingwersen, J., Nowak, W., Gayler, S., & Streck, T. (2024). Within-season crop yield prediction by a multi-model ensemble with integrated data assimilation. Field Crops Research, 308, 109293.

Zeng, Y., Hao, D., Huete, A., Dechant, B., Berry, J., Chen, J. M., ... & Chen, M. (2022). Optical vegetation indices for monitoring terrestrial ecosystems globally. Nature Reviews Earth & Environment, 3(7), 477-493.

Zhang, L., Zhang, Z., Luo, Y., Cao, J., & Tao, F. (2019). Combining optical, fluorescence, thermal satellite, and environmental data to predict county-level maize yield in China using machine learning approaches. Remote Sensing, 12(1), 21.

Zhao, D., Reddy, K. R., Kakani, V. G., Read, J. J., & Koti, S. (2007). Canopy reflectance in cotton for growth assessment and lint yield prediction. European Journal of Agronomy, 26(3), 335-344.

Zhao, Y., Tao, H., He, P., Yao, X., Cheng, T., Zhu, Y., ... & Tian, Y. (2023). Annual 30 m winter wheat yield mapping in the Huang-Huai-Hai plain using crop growth model and long-term satellite images. Computers and Electronics in Agriculture, 214, 108335.

Zheng, Y., Zhang, M., Zhang, X., Zeng, H., & Wu, B. (2016). Mapping winter wheat biomass and yield using time series data blended from PROBA-V 100-and 300-m S1 products. Remote Sensing, 8(10), 824.

Zhuang, H., Zhang, Z., Cheng, F., Han, J., Luo, Y., Zhang, L., ... & Tao, F. (2024). Integrating data assimilation, crop model, and machine learning for winter wheat yield forecasting in the North China Plain. Agricultural and Forest Meteorology, 347, 109909.

Ziliani, M. G., Altaf, M. U., Aragon, B., Houborg, R., Franz, T. E., Lu, Y., ... & McCabe, M. F. (2022). Early season prediction of within-field crop yield variability by assimilating CubeSat data into a crop model. Agricultural and forest meteorology, 313, 108736.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B: Statistical Methodology, 67(2), 301-320.

Appendix A - Supplementary material

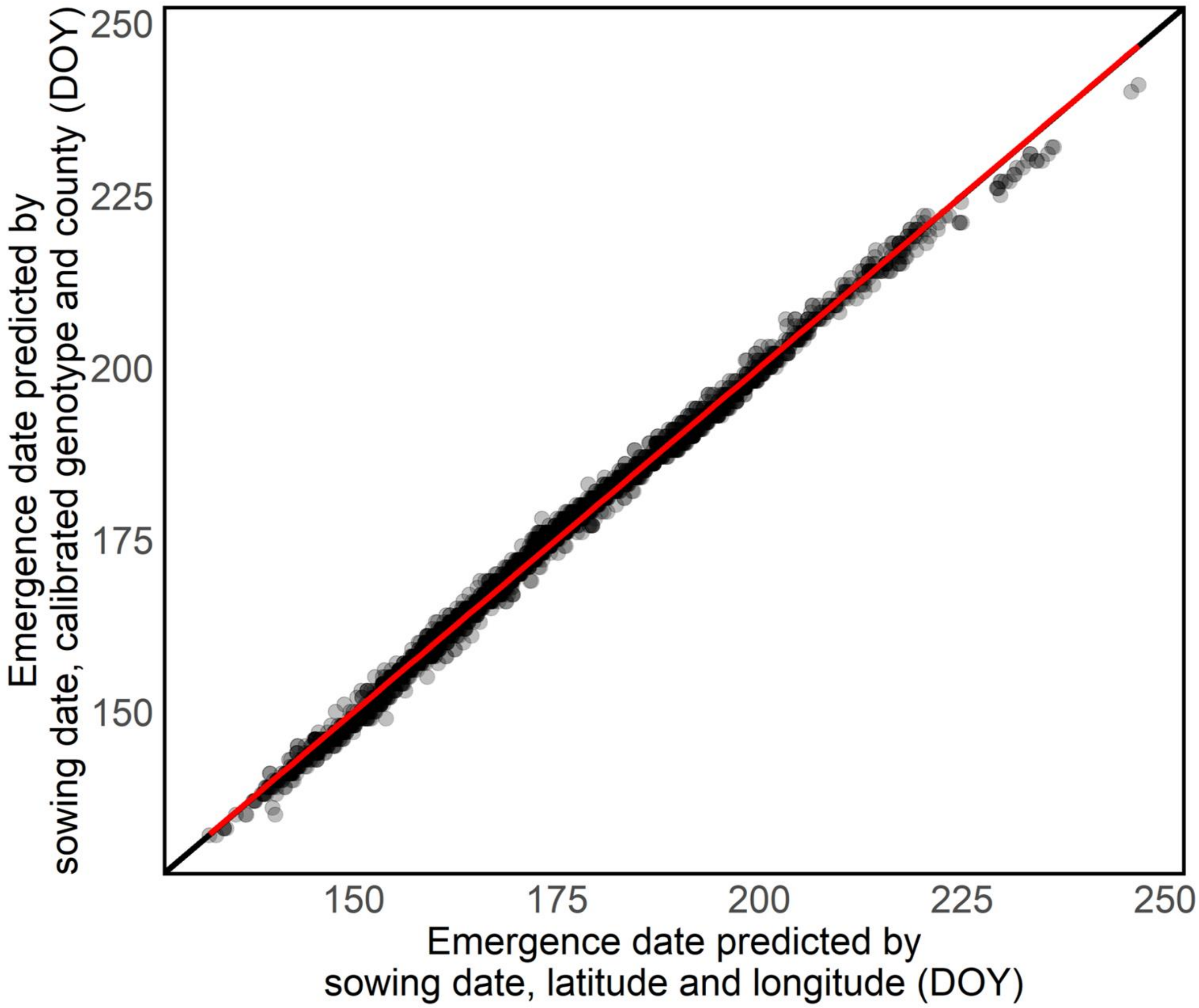

Figure S1: *Emergence date provided by the phenological model 'CronoTrigo' (http://cronotrigo.agro.uba.ar/index.php/) as a function of the modeled emergence date based on sowing date, latitude, and longitude. Each black dot refers to a field-year observation. The black line refers to the 1:1 relationship, while the red line refers to the fit between values on both axes. The x-axis was constructed from the model: Emergence date (DOY) = -37.4 + Latitude * (-0.9) + Longitude * (-0.17) + Sowing date * 1.06. Latitude and longitude are expressed in decimal degrees, and sowing date is expressed as day of the year (DOY). The mean absolute error was 0.81 days.*

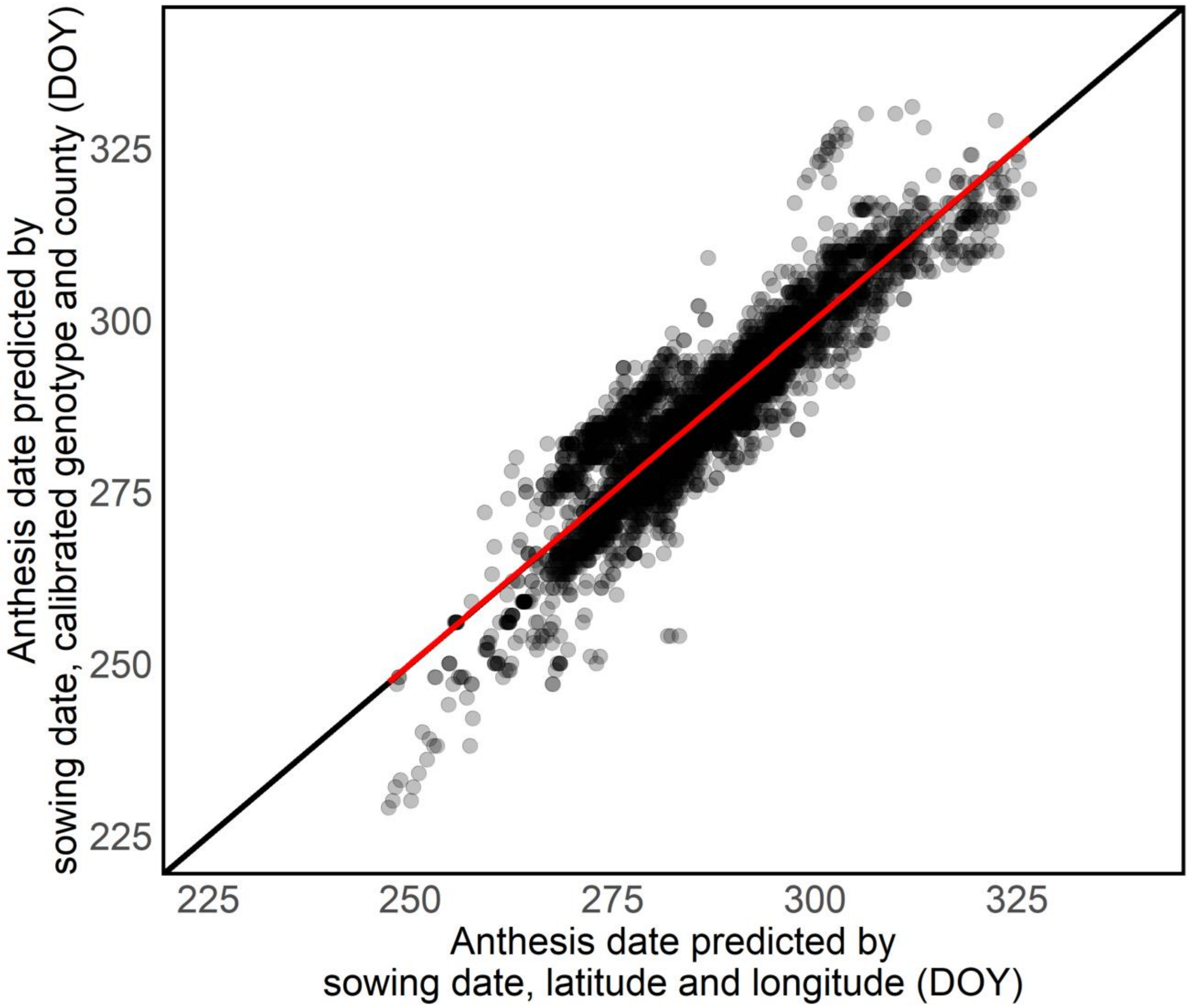

Figure S2: *Anthesis date provided by the phenological model 'CronoTrigo' (http://cronotrigo.agro.uba.ar/index.php/) as a function of the modeled anthesis date based on sowing date, latitude, and longitude. Each black dot refers to a field-year observation. The black line refers to the 1:1 relationship, while the red line refers to the fit between values on both axes. Anthesis Date (DOY) = -610 + Latitude * (-24.5) + Longitude * (-11.1) + Latitude * Longitude * (-0.32) + Sowing Date * 0.30. Latitude and longitude are expressed in decimal degrees, and sowing date is expressed as day of the year (DOY). The mean absolute error was 4.19 days.*

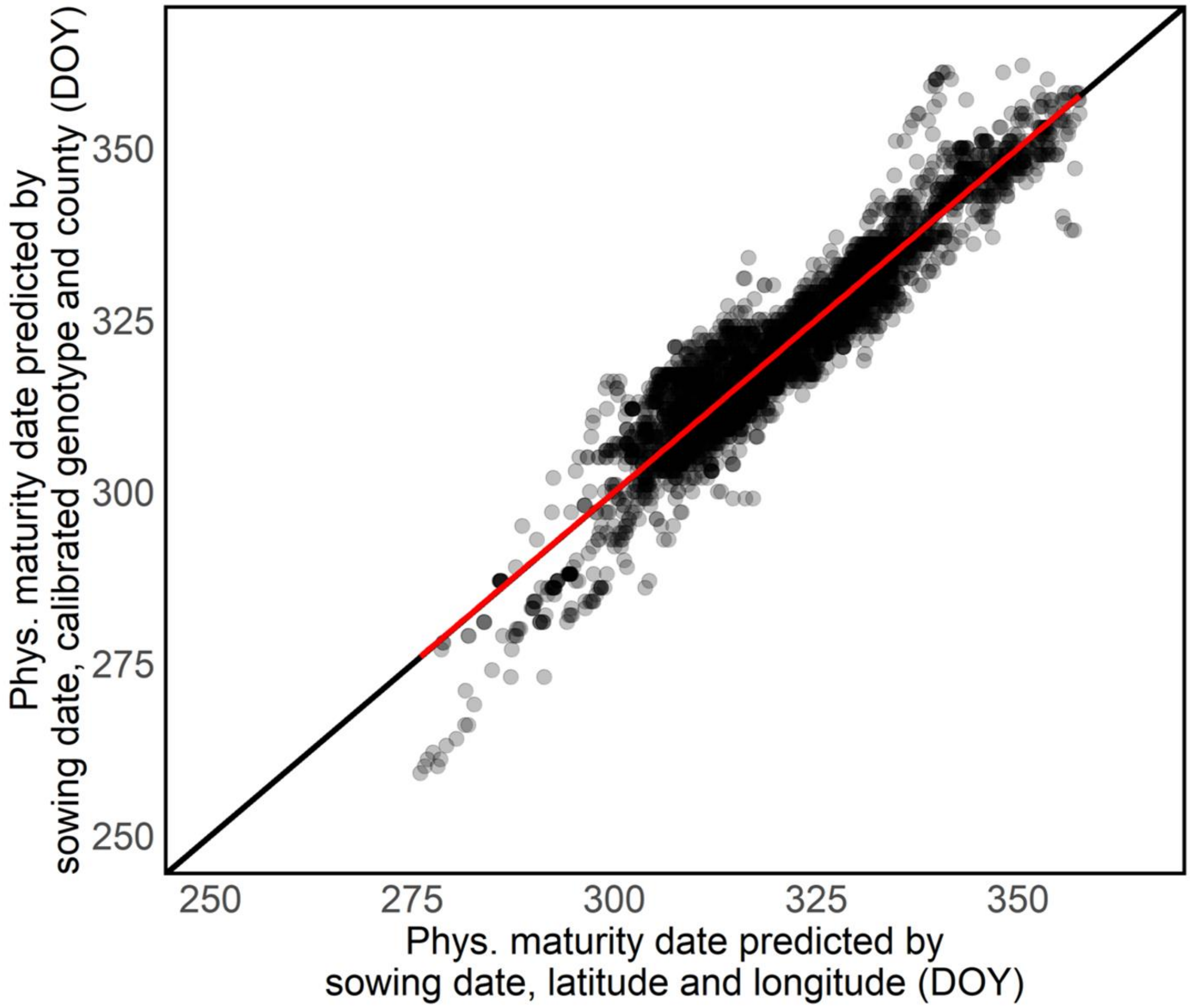

Figure S3: *Physiological maturity date provided by the phenological model 'CronoTrigo' (http://cronotrigo.agro.uba.ar/index.php/) as a function of the modeled physiological maturity date based on sowing date, latitude, and longitude. Each black dot refers to a field-year observation. The black line refers to the 1:1 relationship, while the red line refers to the fit between values on both axes. Physiological Maturity Date (DOY) = 1605 + Latitude * 30.6 + Longitude * 25.3 + Latitude * Longitude * 0.61 + Sowing Date* (-13.9) + Sowing Date * Latitude * (-0.37) + Sowing date * Longitude * (-0.23) + Sowing Date * Longitude * Latitude * (-0.01). The mean absolute error was 3.5 days.*

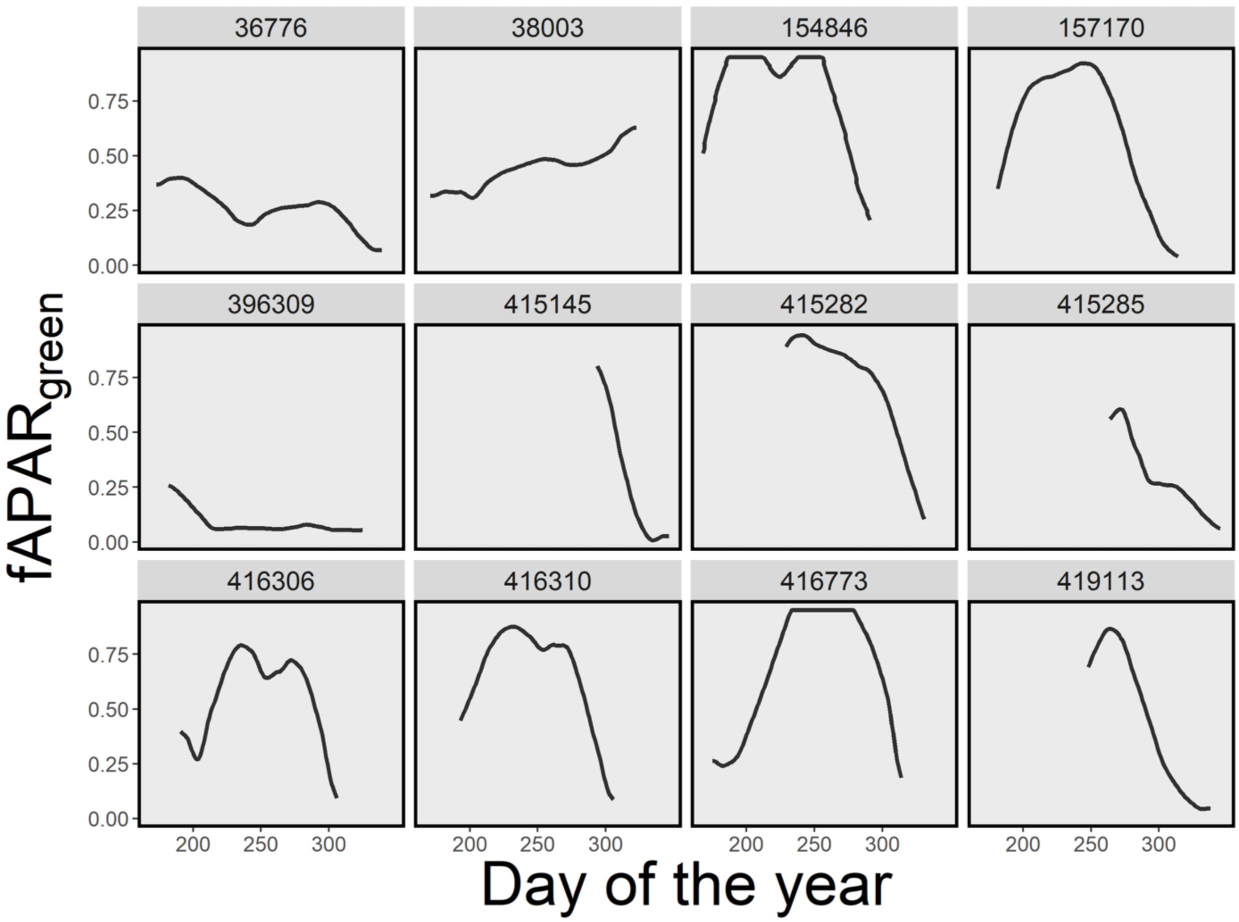

Figure S4: *Example of field-years removed due to an fAPARgreen value greater than 0.25 at sowing. Each panel refers to one field-year. Black lines refer to the smoothed fAPARgreen time series. The start of each black line corresponds to the sowing date reported by the farmer, while the end corresponds to the estimated physiological maturity date.*

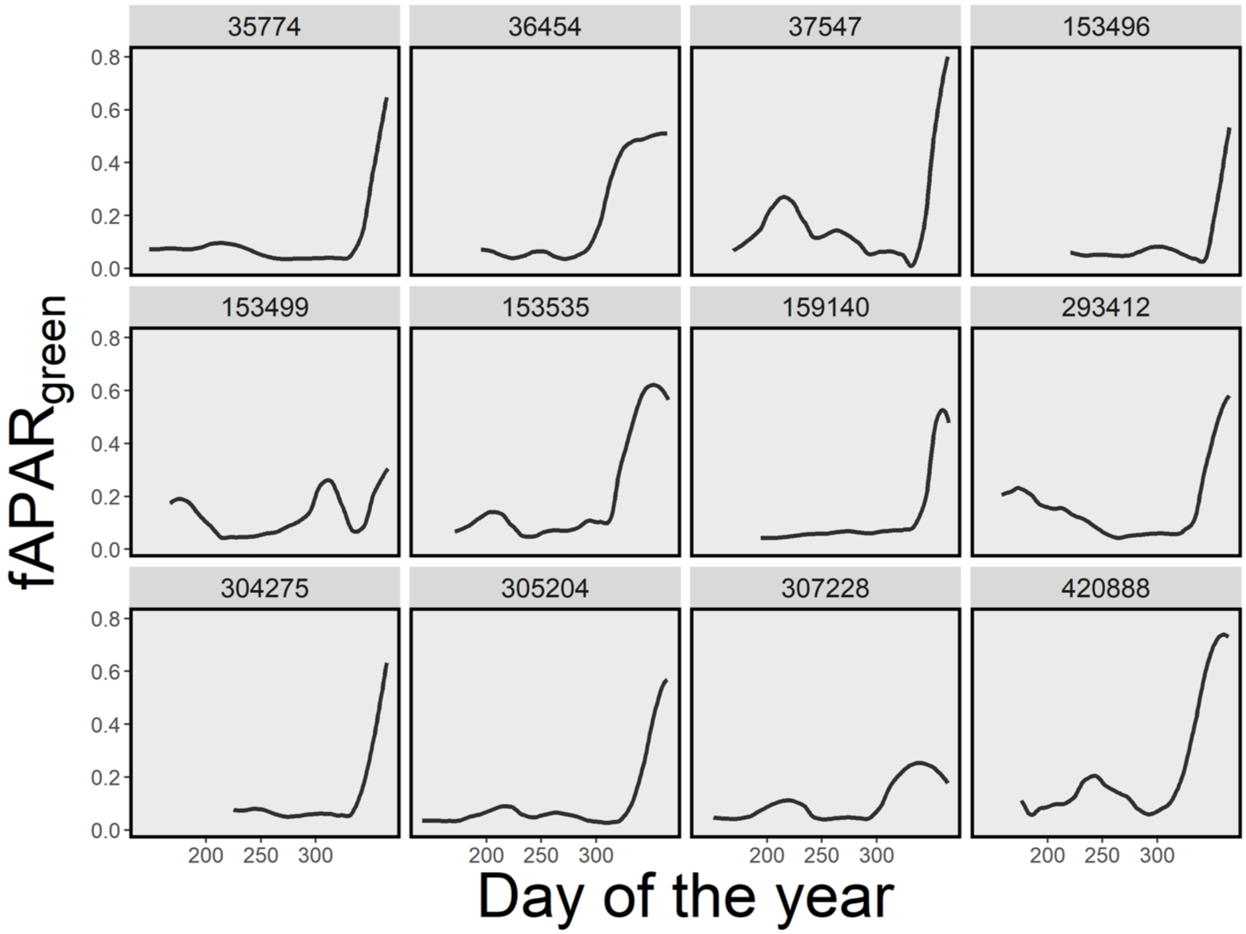

Figure S5: *Example of field-years removed due to a radiation-to-grain conversion efficiency greater than 3 g MJ*$^{-1}$*. Each panel refers to one field-year. Black lines refer to the smoothed fAPARgreen time series. The start of each black line corresponds to the sowing date reported by the farmer, while the end corresponds to the estimated physiological maturity date.*

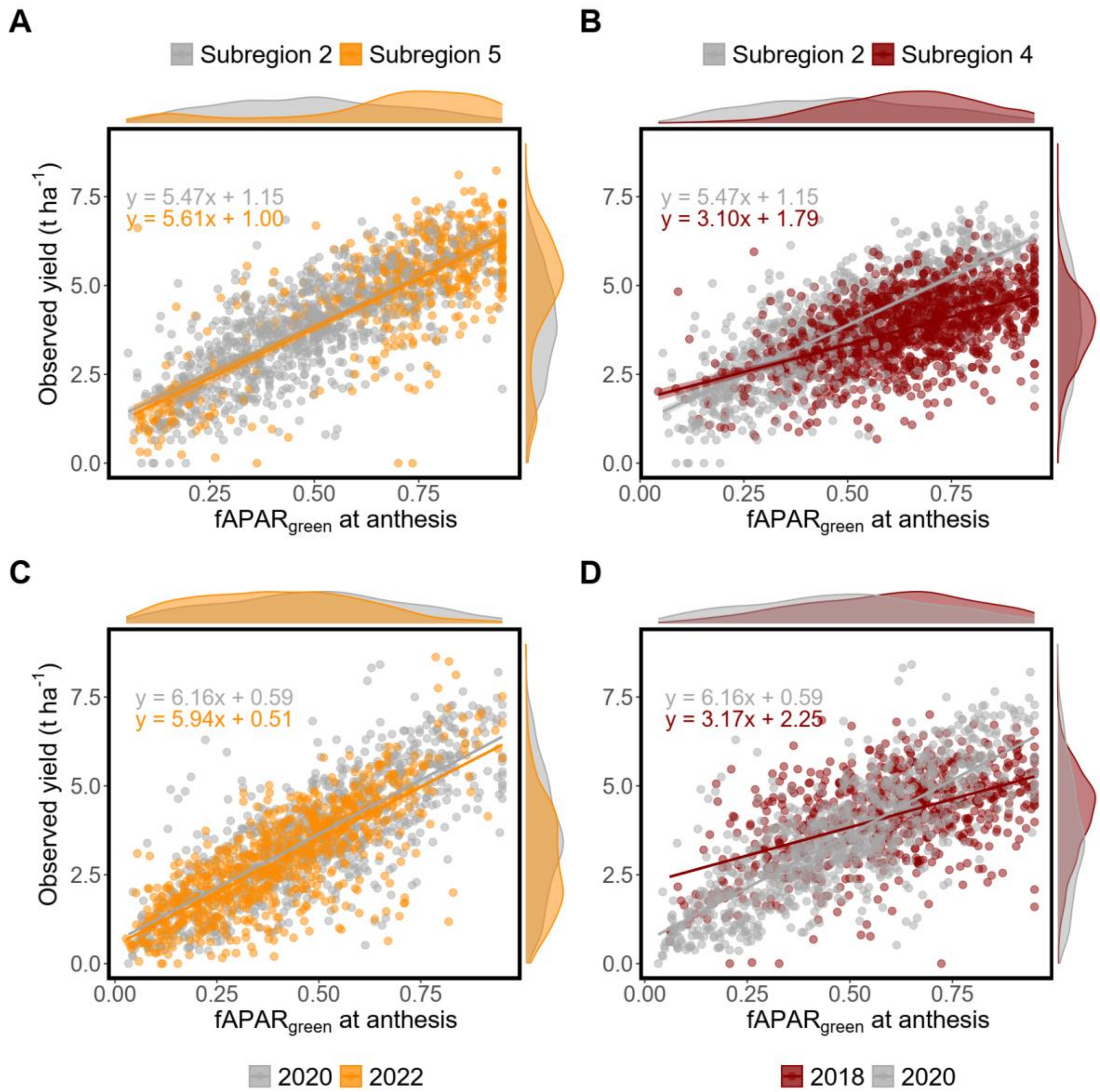

Figure S6: *Observed yield (t ha^{-1}) as a function of the fraction of photosynthetically active radiation absorbed by the crop at anthesis (fAPARgreen). Each point represents a field-year observation, and solid lines indicate the linear fit between axes based on all field-year data. The density plots reflect the distribution of yield and fAPARgreen at anthesis for each database subset. A) Example of two subregions with a similar relationship between yield and fAPARgreen at anthesis. B) Example of two subregions with a different relationship between yield and fAPARgreen at anthesis. C) Same as A, but with years. D) Same as B, but with years.*

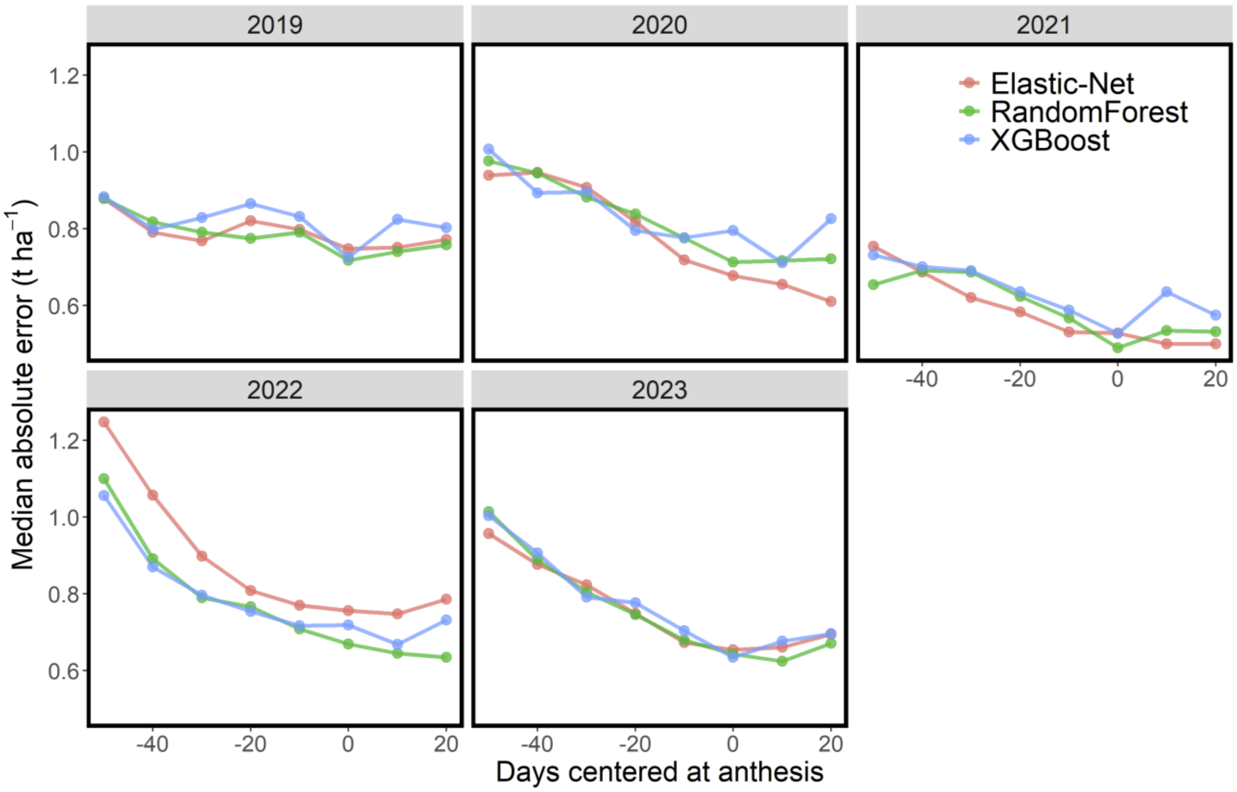

Figure S7: *Median absolute errors (MAE, t ha^{-1}) as a function of days centered at anthesis for Elastic-Net (red line), Random Forest (green line), and XGBoost (blue line). Each panel corresponds to a forecasted year, with models trained exclusively on data from previous years.*

Table S1: *Models for forecasting wheat yield at different moments of the crop cycle trained based on all available data (5.625 field-years). Yield: t ha^{-1}; fAPAR (fraction of photosynthetically active radiation absorbed the crop): from 50 days before to anthesis, its value correspond to the value at the forecast time, while for forecasts after anthesis, its value correspond to the cumulative daily fAPARgreen from anthesis, unitless in both cases; Temp (average daily temperature): °C; IPAR (cumulative incident photosynthetically active radiation): MJ m^{-2}; Prec (cumulative precipitation): mm; VPD (average daily vapor pressure deficit): MPa. All weather variables were summarized from the emergence date to the forecast time. Coefficients were rounded to four decimals.*

Days centered at anthesis	**Model**
-50	Yield = 7.8379 + 3.2104 * fAPAR - 0.4357 * Temp + 0.0007 * Prec - 0.0047 * IPAR + 0.5922 * VPD
-40	Yield = 7.1966 + 3.6511 * fAPAR - 0.4527 * Temp + 0.0014 * Prec - 0.0041 * IPAR + 1.4509 * VPD
-30	Yield = 5.5487 + 4.1183 * fAPAR - 0.3972 * Temp - 0.0014 * IPAR + 1.3523 * VPD
-20	Yield = 3.9769 + 4.52 * fAPAR - 0.3331 * Temp - 0.0002 * Prec + 0.0001 * IPAR + 1.2656 * VPD
-10	Yield = 2.9649 + 4.8478 * fAPAR - 0.2667 * Temp + 0.0007 * IPAR + 1.0126 * VPD
0	Yield = 0.7608 + 5.0575 * fAPAR - 0.1237 * Temp – 0.0002 * Prec + 0.0023 * IPAR + 0.3446 * VPD
10	Yield = - 0.5541 + 0.4698 * fAPAR - 0.0242 * Temp - 0.0007 * Prec + 0.0031 * IPAR -0.3652 * VPD
20	Yield = - 2.0415 + 0.2557 * fAPAR + 0.0777 * Temp - 0.0003 * Prec + 0.0035 * IPAR - 0.8634 * VPD

Appendix B - Elastic-Net hyperparameter optimization

We optimized the hyperparameters of Elastic-Net using 5-fold cross-validation on the training dataset. The optimal combination of alpha and lambda was identified by minimizing the mean absolute error (t ha^{-1}). The values of alpha evaluated were 0, 0.0001, 0.001, 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1, while lambda was explored within the range of 0 to 100 in increments of 0.1. For each combination of these values, a model was trained on four folds and its mean absolute error was calculated on the remaining fold. The combination of alpha and lambda that yielded the lowest mean absolute error across the five folds was selected as the optimal set of hyperparameters.

Appendix C - Random Forest and XGBoost hyperparameter optimization

We optimized the hyperparameters of Random Forest and XGBoost using 5-fold cross-validation on the training dataset to minimize the mean absolute error (t ha^{-1}). For Random Forest, we tuned the number of trees (ntree) at 100, 500, 1000, 1500, and 2000, and the number of variables per tree (mtry) using sequential values from 1 to 4 (the total number of predictors minus one). For XGBoost, we evaluated the number of trees (nrounds) at 50, 200, and 300, the maximum tree depth (max_depth) at 2, 3, 5, 8, and 10, the learning rate (eta) at 0.005, 0.01, and 0.1, and the proportion of features per tree (colsample_bytree) at 0.7 and 1. In both cases, models were trained on four folds and validated on the fifth. Finally, we selected the combination of hyperparameters that minimized the average mean absolute error across the five folds. This procedure was conducted exclusively for the third objective (predicting future years based on the cumulative training of previous years). We used the libraries randomForest (Liaw & Wiener, 2002) and xgboost (Chen & Guestrin, 2016) in R software (R Core Team, 2023) within the caret framework (Kuhn, 2008).