Biosystems Engineering 264 (2026) 104401

Contents lists available at ScienceDirect

Biosystems Engineering

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Maize yield estimation from Sentinel-2 multi-temporal imagery and CANbus data integration: a non-parametric regression approach

G. Stefanescu Miralles [a,b], A. Biglia [a,*], D. Ricauda Aimonino [a], M. Mattetti [c], P. Gay [a], L. Comba [a]

[a] Department of Agricultural, Forest and Food Sciences (DiSAFA) – University of Torino, Largo Paolo Braccini 2, 10095, Grugliasco, TO, Italy
[b] Department of Computer Science (DI) – University of Torino, Via Pessinetto 12, 10149, Torino, TO, Italy
[c] Department of Agricultural and Food Sciences (DiSTAL) – University of Bologna, Viale G. Fanin 50, 40127, Bologna, BO, Italy

ARTICLE INFO

Keywords:
Precision agriculture
Remote sensing
Multi-temporal satellite imagery
Machine learning
Crop yield

ABSTRACT

In precision agriculture, the assessment and estimation of key crop parameters are crucial aspects for the optimisation of input usage and, as an ultimate goal, for the improvement of yield quality and quantity. In this context, a reliable prediction of yield by remotely sensed imagery is an enabling technology for optimisation. In this work, an innovative method for estimating yield in maize cultivation is presented, which exploits multi-temporal and multispectral Sentinel-2 satellite imagery with supervised Machine Learning (ML) techniques. For model training and validation, yield ground truth experimental data from combine harvesters was used, enabling the yield estimation at sub-field scale. The investigation, which was conducted on five case study plots, involved a preliminary comparison of four ML-based algorithms, trained with raw spectral bands. An assessment of the effect of the training dataset on the yield prediction accuracy was then performed. A set of Vegetation Indices (VIs) and Two Band Indices (TBIs) was also considered for this purpose. Finally, a multi-temporal analysis was conducted, in which the temporal evolution of crop spectral data over the maize growing season was exploited using imageries acquired in different epochs. The obtained results proved that an accurate estimation of maize yield can be reached using a Gaussian process regression model, exploiting multi-temporal features directly provided by the raw spectral bands. The model showed a high accuracy in the estimation of maize yield, even when fed with data acquired during only the maize vegetative phase, thus proving its capacity as a prediction tool.

Nomenclature Table

ANOVA	Analysis of Variance
BBCH	Biologische Bundesanstalt, Bundessortenamt, and CHemische Industrie
CANbus	Controller Area Network
EU	European Union
GNSS	Global Navigation Satellite System
GPR	Gaussian Process Regression
MY	Maize Yield
ML	Machine Learning
MLR	Multiple Linear Regression
MRMR	Minimum Redundancy Maximum Relevance
NIR	Near Infrared
PLS	Partial Least Squares
R	Reflectance
RF	Random Forest
RGB	Red-Green-Blue
RMSE	Root Mean Squared Error
SAT	Sentinel-2 satellite imagery
SVM	Supported Vector Machine
SWIR	Short-wave Infrared
TBIs	Two-Band Indices
VIs	Vegetation Indices
a.s.l.	Above sea level
°C	Celsius degree
Hz	Hertz
kg s^{-1}	Kilogram per second
km h^{-1}	Kilometre per hour
m	Metre
mm	Millimetre
t ha^{-1}	Tons per hectare
λ	Spectral band [nm]

* Corresponding author.
E-mail address: alessandro.biglia@unito.it (A. Biglia).

https://doi.org/10.1016/j.biosystemseng.2026.104401
Received 31 October 2024; Received in revised form 16 January 2026; Accepted 22 January 2026
Available online 4 February 2026

1. Introduction

The effects of global warming have been studied over the past decades, and its short-term impacts have been inferred throughout the world (Spinoni et al., 2017). An increase in drought severity has been evidenced in southern European regions (Vicente-Serrano et al., 2014), as well as a trend towards warmer summers (Twardosz et al., 2021). The impacts of such a meteorological trend, together with the demands of a growing population, have dictated huge challenges in the agricultural sector (Sue Wing et al., 2021). Farmers are being pushed to increase their productivity and the efficiency of their practices, and the crop yield is an essential feature (Rosegrant & Cline, 2003). The development of new technologies that allow the automatic monitoring and digitalisation of agricultural activities makes precision agriculture-based strategies effective approaches to face the challenging demands that are currently arising.

Remote sensing plays a relevant role in such a context: the improved capabilities of Earth Observation Systems have made satellite imagery with high spatial resolution, short revisit times and enhanced spectral content available (Kohrs et al., 2014). This has resulted in new ways of exploiting remotely sensed data to monitor crops, such as the prediction of production yields using satellite data (Mokhtari et al., 2018). Indeed, many studies have proved clear correlation patterns between remotely sensed imagery and yields for several crop types. Some examples are the yield estimation of wheat (Engen et al., 2021), tomatoes (Darra et al., 2023), and maize (Schwalbert et al., 2018). However, the aforementioned works were limited to a regional level, and the validation ground truth data used in these studies were mainly aggregated at the farm scale. This could result in these yield assessment tools not being suitable for precision agriculture tasks or having limits in transferability.

Nevertheless, in the last few years, the possibility of estimating yield with satellite imagery, at a sub-field scale, has been investigated considering the spatial resolution of multispectral satellite imagery. This is the case, for example, of the study presented by Li et al. (2022), in which PlanetScope and Sentinel-2 satellite imagery were used to predict maize yields in intercropped smallholder fields. Moreover, Marshall et al. (2022) exploited a Partial Least Squares (PLS) regression to predict the yields of four different crop types, using both Sentinel-2 and PRISMA remotely sensed data. However, the results of those studies were characterised by weak correlations, due to the reduced number of referenced in-field data necessary for the validation phase. These data were generally gathered manually and pointwise during the harvesting stages. Indeed, this lack of validation data is one of the main limiting factors for the adoption of more advanced data driven approaches, such as Machine Learning (ML) tools.

In this context, the exploitation of data acquired by modern agriculture machinery during operation in the field is a game changer. Modern tractors incorporate embedded sensors, such as those integrated into the CANbus network. CANbus is a communication protocol among electronic sub-systems in agricultural machinery, and it is commonly used to record data from sensors embedded in machines, including yield monitors (Stone et al., 2008). This technology gives access to a huge number of in-field data, with high spatial and temporal resolution (Mattetti et al., 2021). Machine data collected from combine harvesters during harvesting allows accurate yield distributions of a crop to be generated at a finer sub-field scale (Sun et al., 2022; Chen et al., 2022; Zhang et al., 2022; Reitz & Kutzbach, 1996). The data within such yield maps represent an unprecedented source of ground truth data, thus paving the way towards the exploitation of supervised ML algorithms, which have been proved to be enhanced by data integration between satellite and ground sources (Mazzia et al., 2020). However, raw data pre-processing is required to make machine-derived data suitable for training ML models (Kosmowski et al., 2021). Recent studies that have followed this approach have shown a close correlation between crop yield and multispectral satellite imagery at the sub-field scale, whenever large datasets were exploited by ML models. Hunt et al. (2019) predicted a wheat yield with a Random Forest (RF) technique, using Sentinel-2 and environmental data, and they obtained an acceptable accuracy of their predictions (a root mean square error (RMSE) of 0.66 tons per hectare (t ha^{-1})). Segarra et al. (2022) presented a method that allowed them to predict the grain cultivation yield, using Supported Vector Machine (SVM) and RF models with Sentinel-2 imagery, and to prove a close correlation among the data (R^2 values of up to 0.89).

In this work, the focus is paid to the yield prediction of maize (*Zea mays*), since it is a major cereal grain that is cultivated throughout the world, and in particular in Europe. Maize production has been estimated at 56 million tons at the European level, over a cultivation area of more than 9 million ha, and an average yield of 6.3 t ha^{-1} (International Production Assessment Division, 2022). These statistics mean that maize is the second most important cereal grain cultivated in the European Union, and it accounts for over 23 % of the EU's total cereal production (European Parliament, 2019). In some countries, such as Italy, the importance of maize in cereal production is even higher, as it was placed in first place, with an approximate yearly production of 6 million tons (41 % of the total production), in 2021 (Statista, 2022). However, few studies have been published till now on the prediction of maize yields, and those that have been published have focused on country/regional yield estimates. Kang et al. (2020) showed that the XGBoost model achieved reasonably good maize yield estimations in the USA, even 10 weeks prior to harvesting. A similar study, at a regional scale in central Europe, reported excellent results, with R^2 values of up to 0.96 when Neural Network models were used (Harsányi et al., 2023). However, the maize yield assessment did not reach a sub-field level of detail in either work, thereby preventing these approaches from being adopted directly in the precision agriculture domain.

All in all, despite the growing availability of high-resolution satellite imagery, reliable methods for estimating maize yield at the sub-field scale remain limited, particularly approaches that effectively integrate multi-temporal and multispectral data with ground-based yield measurements. This study addresses this gap by developing a method for estimating maize yield at the sub-field scale, based on a supervised ML framework that leverages multi-temporal Sentinel-2 imagery in combination with yield data collected directly from combine harvesters. Yield ground truth was obtained through CANbus monitoring during an experimental campaign conducted across five maize fields in northern Italy and used for model training and validation. We hypothesize that the integration of multi-temporal multispectral Sentinel-2 data with supervised ML models can accurately predict maize yield variability at the sub-field level, providing a robust and scalable solution for precision agriculture applications.

The work is structured as follows: (i) a preliminary study was conducted with the aim of selecting the best model for the specific-case from four different models; (ii) an investigation was performed to assess the predictive capability of the selected model, considering different combinations of input datasets computed with feature engineering analysis, and (iii) a multi-temporal approach was then assessed, considering the best dataset-type, through the combined use of images acquired in different periods during the growing season.

2. Materials and methods

2.1. Study area

The case study was conducted over the 2020 season in five maize fields in the Emilia Romagna Region, in the north of Italy, at an altitude of between 30 and 50 m a.s.l. This region is part of the *"Pianura Padana"*, a floodplain that covers the catchment area of the Po River, and it is delimited by the Italian Alps and Pre-Alps to the north and west, by the Northern Apennines to the south and by the Upper Adriatic Sea to the east. The climatic characteristics of the area are typical of a humid

subtropical land (Kottek et al., 2006), due to the confrontation between plains and surrounding mountainous regions, which causes high levels of relative humidity throughout the year. In 2020, the Emilia Romagna region experienced an average rainfall of 822 mm, with most of the rain events being concentrated between June and August (Arpae, 2020). The average temperature in the region was 13.4 °C, and it was the fifth warmest year since 1961.

Maize is one of the major summer cultivations in this area. Its growth and development can be divided into vegetative (from BBCH 07 to BBCH 37) and reproductive stages (from BBCH 52 to BBCH 78) (Meier, 2001). During the vegetative phase, the plant progresses through leaf stages, designated as V1, V2 …, VT, with VT referring to the production of the tassel, which indicates the completion of the total vegetative growth of the plant (Abendroth et al., 2011). The vegetative phase usually lasts around 60 days from emergence, and the reproductive phase follows; this phase starts with silking and pollination and is followed by kernel development. When the kernels are filled to full capacity, a black layer appears at the base of the kernel, and maturity is reached. The reproductive phase extends for a period of approximately two months. Maize is sown in early April in northern Italy, and the peak of the vegetative phase is usually reached around June (Azar et al., 2016). The reproductive phase extends from July to August, and the crop is harvested in September. In the fields selected as case study, seeds are sown during the first weeks of April and are harvested on 9th, 10th, 18th, 28th and 30th of September. The five fields had surfaces of 9.3, 8.1, 7.9, 11.3 and 10.2 ha, respectively, and an average grain yield of 9.3 t ha^{-1}. In the following sections, the five fields are referred to with the letters “A”, “B”, “C”, “D” and “E”.

2.2. Multispectral satellite imagery

Sentinel-2 satellite imagery (SAT) is selected for this work for the following reasons: (i) it offers a spatial resolution of up to 10 m and a revisit time of 2/3 days at mid-latitudes (Defourny et al., 2019), (ii) it is available free of charge, thereby boosting the distribution and commercialisation of remote sensing solutions based on this satellite mission, and (iii) the type and number of spectral bands available have proved to be effective in many agricultural applications. The Sentinel 2 spectral bands involve 4 red-edge and near infrared (NIR) narrow bands (<20 nm), which, being sensitive to changes in chlorophyll content and canopy structures, represent potential predictors of grain yield (Frampton et al., 2013).

A set of satellite multispectral images, acquired on 12 different dates throughout the maize growing cycle (see Fig. 1), is downloaded from the Copernicus Open Access Hub (Copernicus European Union ©, 2023) for each of the five fields, and the images are here referred to as SAT_i, with $i = 1, \ldots, 12$. Only satellite images from four weeks post-sowing until harvesting are selected, and, in addition, any tiles presenting a cloud coverage above 7 % are discarded. The SAT_i dataset is georeferenced in the UTM/WGS84 projection, with a bottom of atmosphere reflectance correction. Each satellite multispectral image comprises 10 spectral bands, the main features of which are shown in Table 1.

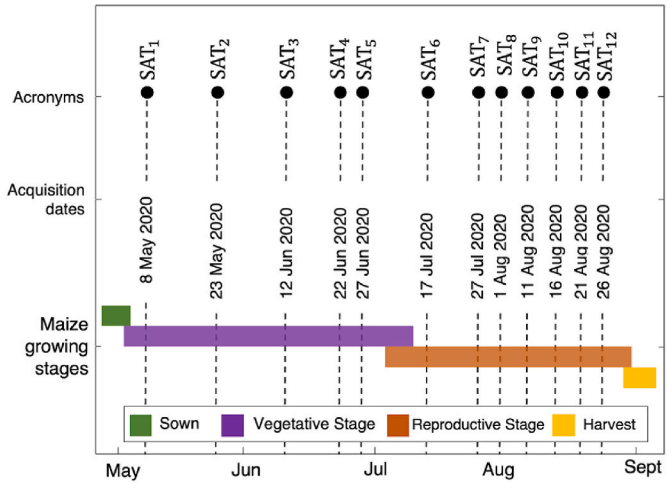

Fig. 1. Selected Sentinel-2 imagery acronyms together with their acquisition dates and their location within the maize life cycle.

2.3. Reference maize yield map

The maize yield distribution maps of the five fields are computed using the information conveyed from a monitored combine harvester with the CANbus protocol (Stone et al., 2008). A New Holland CR7.90 (New Holland, Zedelgem, Belgium), with a header bar width (w) of 7.62 m, is selected for this work. The harvester is equipped with an impact-based yield monitor system, which provided the harvested crop flow rate in kg s^{-1}. A stand-alone, CANbus datalogger (Kvaser Memorator pro 2xhs v2, HQ Kvaser AB, Sweden) is used to record the working parameters of the machine. The CANbus datalogger has an embedded global navigation satellite system (GNSS) receiver that allows the harvester velocity to be recorded in km h^{-1} as well as the position of the vehicle during in-field operation. The datalogger has an acquisition frequency of 10 Hz and a probable circular error of 2.5 m. The yield monitor is calibrated before harvesting, using a single-point calibration approach (Schuster, 2016). The combine harvester is driven by a professional driver with more than 20 years of experience.

The crop flow and the machine velocity, together with the header width, are used to calculate the local maize yield associated with each GNSS position. In this phase, the time-lag between the position of the harvester and crop flow that appeared, due to the harvester dynamics (Chung et al., 2002; Yang et al., 2002), is managed as described in Lyle et al. (2014). A rasterisation phase is then applied for each pixel of the SAT dataset. The respective raster maize yield values are computed from the subset of vectorial data included in the pixel area, through an averaging approach that is equivalent to the point-to-raster method in the ArcGIS Spatial Analysis Menu Tool (ArcGIS Pro, 2017). The rasterisation process results in a total of 7454 pixels covering the five fields. Finally, a Gaussian filter is applied to the maize yield field maps as a smoothing step, according to the approach presented by Wen et al. (2021). The maize yield frequency distributions of the five fields “A”, “B”, “C”, “D” and “E” are displayed in Fig. 2, together with the overall distribution of the whole merged dataset.

2.4. Complete dataset overview

In summary, the reference ground truth Maize Yield (MY) map of the five fields had 7454 pixels. The corresponding spectral values (10) in the

Table 1
Sentinel-2 multispectral bands used in this study, with their corresponding central wavelenght, bandwidth and spatial resolution.

Spectral Band	Central Wavelength [nm]	Spatial Resolution [m]	Bandwidth [nm]
B02: Blue	490	10	65
B03: Green	560	10	35
B04: Red	665	10	30
B05: Red edge 1	705	20	15
B06: Red edge 2	740	20	15
B07: Red edge 3	783	20	20
B08: NIR 1	842	10	115
B8A: NIR 2	865	20	20
B11: SWIR 1	1610	20	90
B12: SWIR 2	2190	20	180

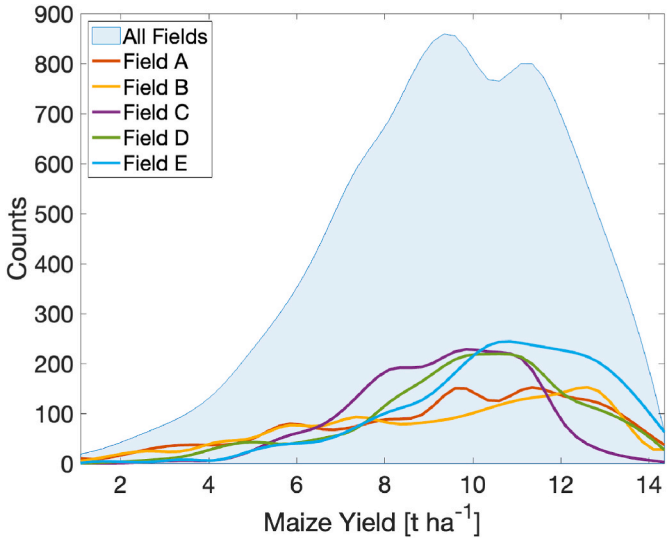

Fig. 2. Frequency distribution of ground-truth maize yield and its cumulative curve. The curves have been derived from combine harverster CANbus data acquired during the harvesting phase of the case study fields.

SAT_i datasets are selected for each MY pixel value. This procedure is repeated for the 12 available dates, and this led to approximately 890,000 discrete values. As an example, the vectorial GNSS position of the harvesting path followed by combine harvesters is shown in Fig. 4, acquired on 18th, 28th and 30th September 2020 in three of the five fields "C", "D" and "E". The rasterised MY map is also shown in Fig. 4. The lower layer of the subplots represents the Red-Green-Blue (RGB) visualisation of the multispectral satellite images acquired on 11th August 2020

2.5. Methodology phases

In this study, a methodology to achieve high-accurate maize yield estimates from remotely sensed Sentinel-2 multispectral imagery is developed. The analysis can be divided into three main phases: (1) a *Machine Learning model selection* phase, in which the performance of four different ML models is evaluated and compared, (2) a *Feature Engineering* phase, in which the training input dataset that preserved the maximum informative content of the crop status is selected, and (3) the final phase, in which the MY prediction is improved by exploiting the temporal evolution of the spectral response of the crop during the growth cycle, as presented in the *Multi-temporal analysis* section. The same performance metrics, which are described in the last part of the Methods section, are adopted for the entire investigation. Fig. 3 shows a work-flow schematic for the described processes.

2.6. Machine learning model selection

A performance comparison between different ML models is carried out to develop an accurate approach to maize yield estimation from satellite imagery. Specifically, according to results presented in the literature, the following ML models are selected: Random Forest (RF) (Everingham et al., 2016; Hunt et al., 2019), Multiple Linear Regression (MLR) (Marshall et al., 2022; Qader et al., 2023), Supported Vector Machine (SVM) (Oguntunde et al., 2017; Saruta et al., 2013; Segarra et al., 2022) and Gaussian Process Regression (GPR) (Ghosh et al., 2022). The performance of the ML models is assessed using the same dataset type in the form of raw spectral bands in mono-temporal format, each consisting of approximately 7500 pixels. A k-fold cross-validation approach (with $k = 5$) is adopted as a robust method to derive independent dataset performance scores. The comparison of the ML models is repeated using satellite data from six different dates, selected from the

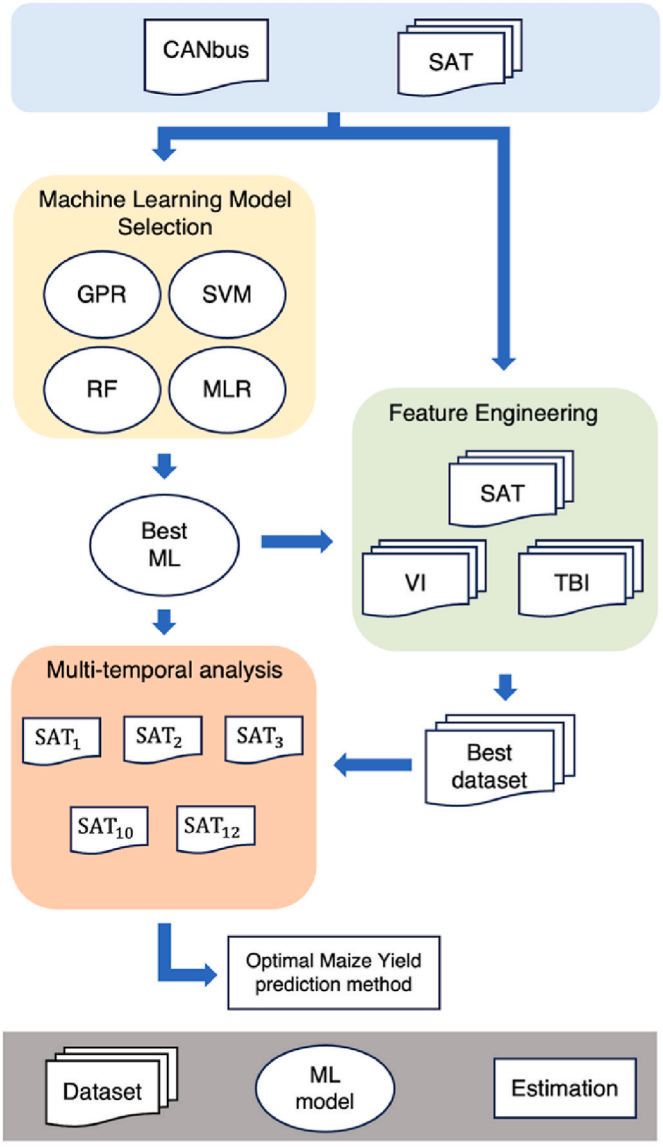

Fig. 3. Work-flow scheme of the different processes and analysis carried out in the work.

maize life cycle, to verify that the best performing detected model is always the same, independent of the acquisition date. In this phase, the used datasets are: SAT_1 and SAT_2 in May (vegetative growth, leaf and cob development), SAT_3 and SAT_4 in June (pollination and kernel development), SAT_6 in mid-July (grain filling), and SAT_{10} in mid-August (grain filling/physiological maturity).

Details of each of the four ML models and the adopted parameter settings are reported hereafter.

1) In an RF algorithm (Breiman, 2001), input training data are randomly resampled to create a specific number of smaller sets, which are then fitted to individual decision trees. Each tree of the ensemble uses a different subset of explanatory variables to split the data at each node, thus generating different response predictions. The estimated value of a continuous response is then computed as the mean of all the individuals.
2) The MLR approach consists of reducing a matrix of features (X) to an additive series of factors that explain a linear proportion with a response variable (Y). The new factors are the result of a linear transformation, in the normalised direction, that maximises the X-Y correlation (Atzberger et al., 2010), and each factor is characterised

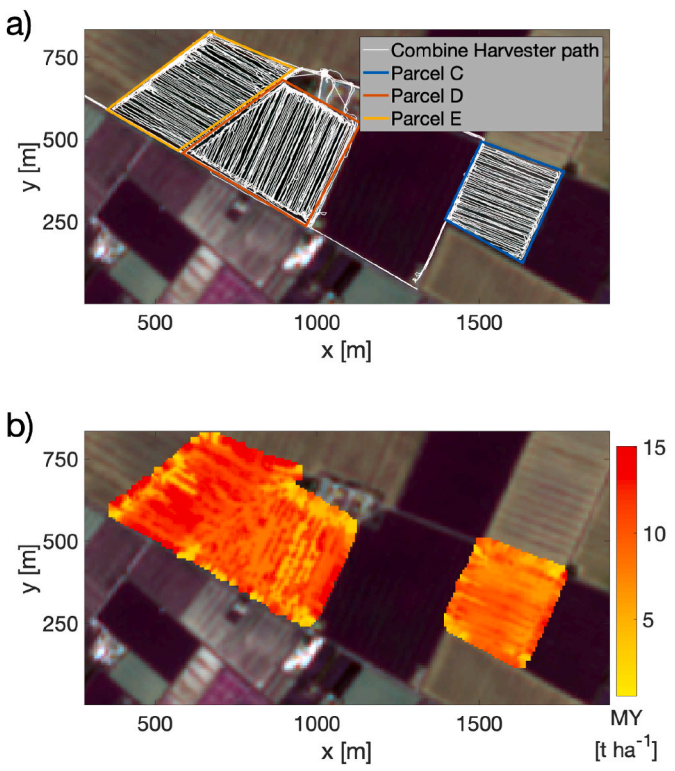

Fig. 4. RGB visualisation of the multispectral satellite imagery acquired on 11th August 2020, showing three of the five experimental fields condidered. (a) Vetorial GNSS position of the combine harvester path, acquired on 18th, 28th and September 30, 2020, and (b) MY ground truth in false colour, resulting from the rastering procedure. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

by a weight that describes its relative contribution to the response variable.

3) SVM is a well-known methodology that was first developed by Boser, Guyon, and Vapnik (1992) to approach multidimensional classification problems. However, SVM has also been adapted to regression problems (Cristianini & Shawe-Taylor, 2000). The idea behind such a method is finding a good data-fitting hyperplane in the feature space, by means of the optimisation of a kernel function. The resulting hyperplane equation can then be used to predict response variables from a new set of features.
4) GPR is a non-parametric probabilistic method used for regression that is based on the assumption that the studied data can be described by a Gaussian process (Rasmussen, 2004). A Gaussian process is defined as a joint distribution of random variables that follow a normal distribution. A prior distribution, from which the training data is used to update the parameters, is assumed from among all the possible sets of functions that a Gaussian process can take, and this process concludes when the most suitable function to fit the data is obtained. Once the training has been completed, the joint distribution of the training target values can be retrieved and then used to predict unseen observations with an associated uncertainty (Ghosh et al., 2022).

The ML model assessment and selection steps are implemented in the Matlab® R2023 Regression Learner app (MathWorks®, 2024) for all four models described previously. For reproducibility, details of parameter settings are reported below: the RF algorithm is configured as an ensemble of 100 boosted decision trees, with a minimum leaf size of 8 and a learning rate of 0.1; the MLR model does not require hyperparameter specification, as only linear terms are considered; the SVM and GPR models, however, require the definition of kernel functions, which have been selected as quadratic and Matern 5/2 kernels, respectively. The differences between the performances of the ML model are statistically verified by an ANOVA test on the obtained metric scores.

2.7. Feature engineering

To optimally exploit the informative content provided by the SAT spectral data, a feature engineering approach is followed to determine the most effective set of inputs for the ML model. In addition to the *SAT* raw spectral bands, a set of vegetation indices (VIs) and a set of Two-Band Indices (TBIs) are considered for this task. The set of VIs is selected on the basis of the state-of-the-art literature works (Hunt et al., 2019; Kanke et al., 2016; Kaya & Polat, 2023; Panda et al., 2010; Satir & Berberoglu, 2016), and their definitions are reported in Table 2. The VIs are calculated from two or more spectral bands of the SAT tiles. The definitions of the set of TBIs followed the approach proposed by Marshall et al. (2022): the TBIs are computed from all the possible combinations of two spectral bands, λ_a and λ_b, using the normalised difference equation:

$$TBI_{\lambda_a,\lambda_b} = \frac{R_{\lambda a} - R_{\lambda b}}{R_{\lambda a} + R_{\lambda b}} \quad \text{Eq. (1)}$$

where $R_{\lambda 1}$ and $R_{\lambda 2}$ are the reflectances measured in the two spectral bands. As SAT has 10 spectral bands 36 $TBI_{\lambda_a,\lambda_b}$ combinations are possible. Three sets of data are thus considered:

- 10 *SAT* raw spectral bands (SAT);
- 8 vegetation indices (VIs);
- 36 TBIs.

Three feature selection strategies are implemented on each of the three sets (SAT, VIs, TBIs) to improve the prediction performance and provide faster and more computation-effective predictors (Guyon et al.,

Table 2
Vegetation indices used in the study, with the respective formulas computed from the Sentinel-2 multispectral bands.

Vegetation Index	Acronym	Formula	Reference
Green Chlorophyl Vegetation Index	GCVI	$\frac{B8A}{B3} - 1$	Gitelson et al. (2003)
Green Normalised Difference Vegetation Index	GNDVI	$\frac{B8 - B3}{B8 + B3}$	Gitelson et al. (1996)
Normalised Difference Vegetation Index	NDVI	$\frac{B8 - B4}{B8 + B4}$	Rouse et al. (1973)
Simple Ratio	SR	$\frac{B8}{B4}$	Jordan (1969)
Water Dynamic Range Vegetation Index	WDRVI	$\frac{0.2*B8A - B4}{0.2*B8A + B4}$	Gitelson (2004)
Enhanced Vegetation Index	EVI	$2.5\left(\frac{B8 - B4}{0.2*B8 - 6*B4 - 7.5*B2 + 1}\right)$	Haboudane et al. (2002)
Triangular Greenness Index	TGI	$-\frac{1}{2}[190(B4 - B3) - 120(B4 - B2)]$	Hunt Jr et al. (2013)
Normalised Green Red Difference Index	NGRDI	$\frac{B3 - B4}{B3 + B4}$	Hunt et al. (2005)

2003): (1) the RReliefF algorithm (Robnik-Šikonja & Kononenko, 2003), (2) the Minimum Redundancy Maximum Relevance (MRMR) algorithm (Ding & Peng, 2005), and (3) the F-test (Sureiman & Mangera, 2020). In this work, the built-in features selection algorithms implemented in the MATLAB programming environment are used, using their default parameters settings. Overall, to determine the most effective set of inputs for the ML model, 12 dataset types are thus considered, as detailly described in Table 3. The effectiveness of the sets of features is investigated considering data from all the 12 acquisition dates, using a mono-temporal approach, leading to an overall set of 144 set of features (12 types for 12 dates).

As in the ML model selection process, a k-fold cross-validation approach with $k = 5$ is adopted to derive representative performance scores for each of the sets of features. Only a single ML model, that is, the best one obtained from the approach described in the previous section, is used in this section.

2.8. Multi-temporal analysis

The potential of a joint exploitation of data from different maize phenological phases is investigated using the best combination of input-features and the best ML model determined in the prior steps. To this aim, multi-temporal datasets, generated by combining groups of single-date datasets, are considered. More specifically, groups of 2–5 single-date datasets are investigated. Each multi-temporal dataset is obtained by means of a concatenation of the features from individual acquisition dates. For example, if two datasets from single acquisition dates made up of 10 features each are considered, a bi-temporal dataset that jointly exploits them would be made up of 20 features. From the 12 available dates, only a few single-date datasets are used, that is, the best ones from the *Feature Engineering* phase. Again, in this phase, the performance of the ML model fed with multi-temporal datasets is validated by means of the k-fold cross-validation method, with $k = 5$.

2.9. Metrics

Two accuracy metrics are adopted to compare the performance of the several tested models and input datasets: the coefficient of determination R^2 and the Root Mean Squared Error (RMSE). R^2 is computed as

Table 3
Definition of the 12 considered dataset types used as input of the ML model in the feautre engineering phase.

Feature Type	Feature selection method	Nomenclature	Description
SAT	none	SAT_All	All the 10 SAT raw spectral bands
	RReliefF	SAT_RReliefF	Top 4 SAT raw spectral bands select with RReliefF algorithm
	MRMR	SAT_MRMR	Top 4 SAT raw spectral bands select with MRMR algorithm
	F-test	SAT_Ftest	Top 4 SAT raw spectral bands select with F-test algorithm
VIs	none	VIs_All	All the 8 VIs defined in Table 2
	RReliefF	VIs_RReliefF	Top 4 VIs select with RReliefF algorithm
	MRMR	VIs_MRMR	Top 4 VIs select with MRMR algorithm
	F-test	VIs_Ftest	Top 4 VIs select with F-test algorithm
TBIs	none	TBIs_All	All the 45 TBIs
	RReliefF	TBIs_RReliefF	Top 10 TBIs select with RReliefF algorithm
	MRMR	TBIs_MRMR	Top 10 TBIs select with MRMR algorithm
	F-test	TBIs_Ftest	Top 10 TBIs select with F-test algorithm

$$R^2 = 1 - \sum_{i=1}^{N} \frac{(MY_{Est\,i} - MY_{Obs\,i})^2}{(MY_{Obs\,i} - \underline{MY_{Obs}})^2} \quad \text{Eq. (2)}$$

where MY_{Est} and MY_{Obs} are the N element vectors of the estimated and observed values of the variable being predicted, and $\underline{MY_{Obs}}$ represents the mean of MY_{Obs} as

$$\underline{MY_{Obs}} = \frac{1}{N} \sum_{i=1}^{N} MY_{Obs\,i} \quad \text{Eq. (3)}$$

The RMSE, which is related to R^2, is computed as

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (MY_{Est\,i} - MY_{Obs\,i})^2} \quad \text{Eq. (4)}$$

in order to provide a dimensional indicator of the model accuracy.

The selection of R^2 and RMSE as the primary metrics for evaluating the ML methods examined in this study is based on a comparative assessment of related literature in this research area. Using equivalent evaluation metrics in this work enables straightforward comparisons across studies and facilitates clearer interpretation of the improvements reported.

3. Results and discussion

3.1. Machine learning model selection

The accuracy of the GPR model in predicting MY was found to be much higher than that of the other considered models. For each model-dataset pair (4 tested models trained with 6 datasets from different dates), the mean and standard deviation of obtained R^2, resulting from the 5-fold cross validation, are reported in Fig. 5a. In addition, all the obtained R^2 values are graphically represented in the boxplot in Fig. 5b, grouped by ML model. Considering the mean metric per model, the GPR model showed a mean R^2 of 0.66, clearly outperforming the accuracies reached by the other models. The results, in terms of average RMSE (1.50 t ha^{-1} for GPR, 2.10 t ha^{-1} for SVM, 2.06 t ha^{-1} for RF and 2.24 t ha^{-1} for MLR), showed a reduction in the error obtained by the GPR model of more than 0.6 t ha^{-1}, when compared to the other models. Moreover, the superiority of GPR was verified for every date, as it showed an independence of this result from the acquisition date of the input data.

This result agreed with those of Ghosh et al. (2022), who estimated wheat biophysical variables from satellite imagery using different ML models (GPR, SVM and RF) and found GPR was the model that provided the best outcomes. Also, in agreement with Segarra et al. (2022), it was observed that the usage of ML-based approaches resulted in better estimates than a linear regression model, thus suggesting a more complex correlation between crop yield and reflectance, which a GPR can exploit. However, Segarra et al. (2022) observed that RF outperformed SVM, while these two models in this study showed equivalent accuracies. In the work of Marshall et al. (2022), the RF model also provided better results than a PLS method, in which PRISMA and Sentinel-2 images were used to estimate the yield of four different types of crops (maize, rice, soy and wheat). Other authors have reported unclear differences between simple linear regression and other ML models, such as Artificial Neural Networks or Random Forests (Uno et al., 2005; Zhou et al., 2020). However, the presented results here showed that the selection of the ML model was a crucial phase, since the performance of the model on the same type of data can differ notably for different ML algorithms. Thus, to achieve a high accuracy in the estimation of the maize yield with Sentinel-2 imagery, the GPR model was selected and used as the reference model in the subsequent investigation phases.

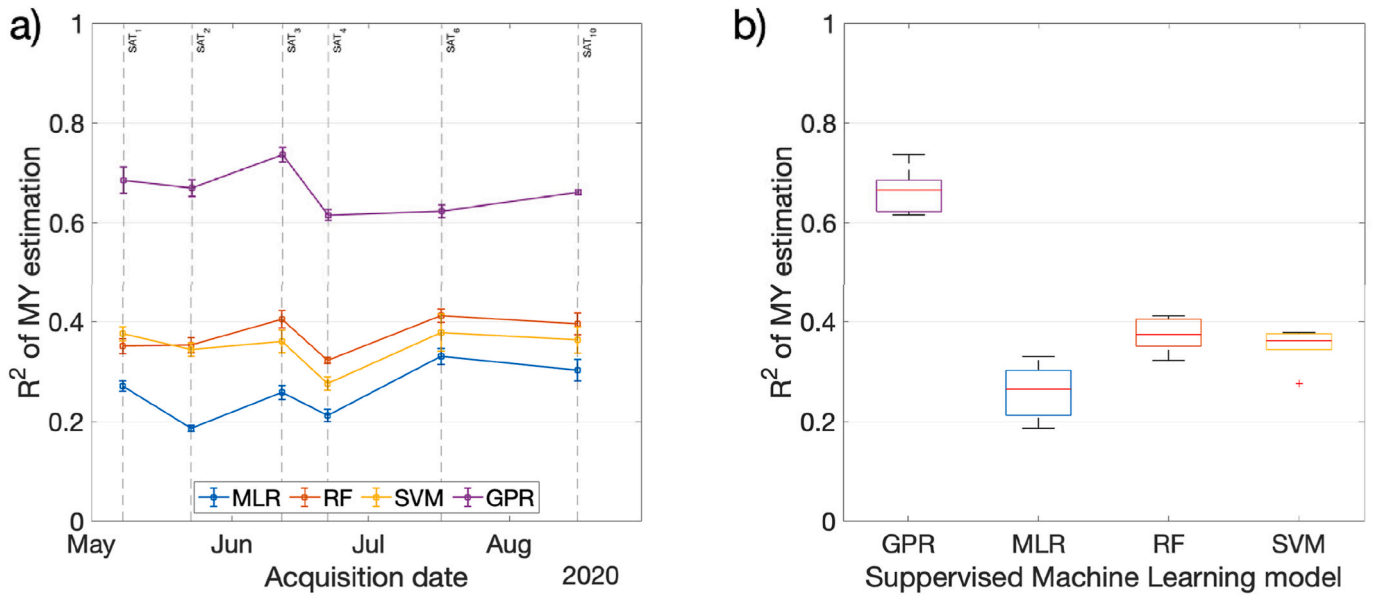

Fig. 5. (a) MY estimation R^2 score of each model-dataset pair assessment, with the k-fold cross validation results represented as the average (square) and standard deviation (error bar). (b) Boxplot of R^2 score grouped by ML model.

3.2. Feature engineering results

The *Feature Engineering* investigation focused on determining the most effective set of inputs for the ML model. It should be recalled that the considered three input-types of data were Sentinel 2 raw spectral bands (*SAT*), a set of vegetation indices (VIs), and a set of Two-Band Indices (TBIs). The first assessment centred on studying differences within subsets of the same input-type data. This was aimed at checking the efficiency of the subset selection on the basis of feature engineering strategies. The selected features of each SAT_i tile are displayed in Fig. 6, for the three dataset types and the three feature selection algorithms (RReliefF, MRMR and Ftest).

The best MY estimates of the SAT approach corresponded to the entire combination of all 10 spectral bands, instead of a combination of subsets from feature selection strategies (RReliefF, MRMR, Ftest). The MY estimation results, when the ML model was fed with SAT_All, showed R^2 values higher than when it was fed with SAT_RReliefF, SAT_MRMR and SAT_Ftest, respectively, as depicted graphically in Fig. 7 in yellow lines. Results from the RReliefF and MRMR algorithms showed similar performances, with Ftest being the feature selection algorithm that provided the poorest outcomes. As far as the feature selection strategies were concerned, it was noted that the RReliefF algorithm showed a high robustness in the choice of its SAT bands. The RReliefF algorithm always selected both the B8A and the B11 bands over the 12 analysed dates. B8A corresponds to an NIR band, with wavelengths ranging from 855 to 875 nm, while the B11 band belongs to the Short-wave Infrared (SWIR) range of the spectra, that is, from 1565 to 1655 nm. The B05 band (698–712 nm), which corresponds to the red edge, was selected 10 times, while B06, which is also in the red edge (733–747 nm), was also selected 10 times, mostly on phenological maturity dates. None of the bands corresponding to the visible part of the spectra were selected (B02, B03 or B04), thus indicating that the SAT bands that contained the most information that could be correlated with yield were in the red edge, NIR and SWIR zones. Kang et al. (2020) was in agreement with the RReliefF algorithm selection criteria and stated that from among all the considered bands from the MODIS satellite, the NIR band showed the closest correlation with maize yield. Chauhan et al. (2020) observed a similar trend in a study centred on identifying wheat lodging, where the Sentinel-2 red edge and NIR bands were the best at distinguishing healthy wheat from lodged wheat. In general, the results

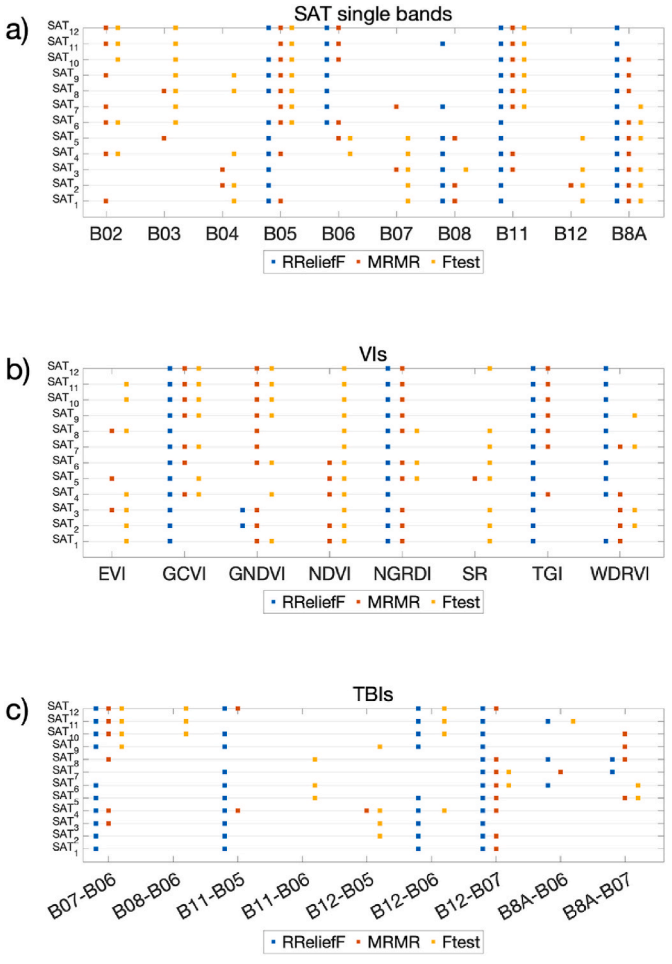

Fig. 6. (a) SAT, (b)VI and (c) TBI selected features corresponding to each SAT_i tile. As indicated in the legend, the different colors are referred to the three different feature selection algorithms (RReliefF, MRMR and Ftest). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

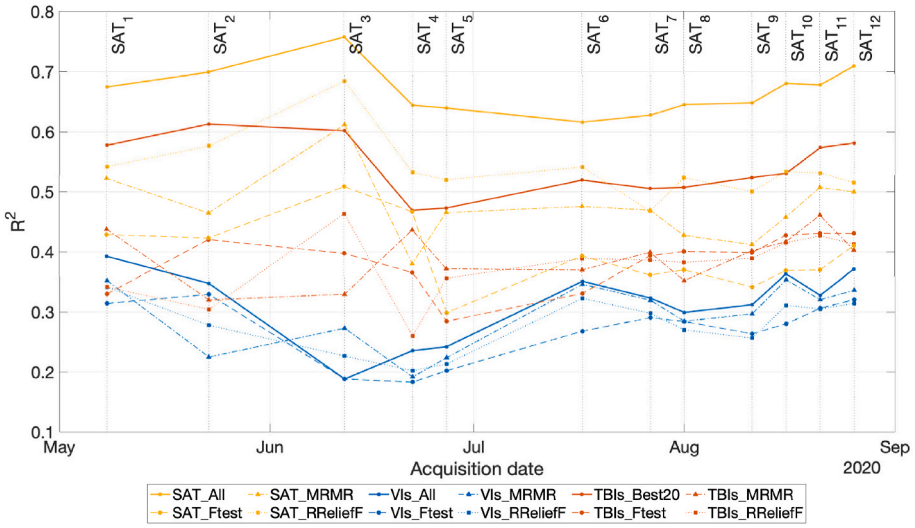

Fig. 7. For the 12 imagery dates, MY prediction results (R^2 score) of GPR model fed with the different 12 input-type datasets described in Table 3.

obtained from the MRMR algorithm agreed with the observations from RReliefF, particularly concerning the B8A, B11 and B05 bands. The selections made by the Ftest algorithm, which resulted in the lowest MY estimations, showed unclear patterns, and only preferences for the B03, B05 and B11 bands in the second half of the maize life cycle. Thus, the results from the SAT datasets suggested that all 10 bands contain valuable information for MY estimation, and that the GPR model was able to return better estimates when fed with the entire combination of all 10 bands instead of with a smaller subset. Nevertheless, bands corresponding to the red edge, NIR and SWIR were systematically preferred by the feature selection algorithms, indicating a higher content of potential information that could be used for yield estimates than bands in the visible part of the spectrum.

The VIs_All dataset, considering all the VIs defined in Table 2, performed slightly better than the subsets that arose from the feature selection strategies, with R^2 values that were on average higher than the VIs_RReliefF, VIs_MRMR and VIs_Ftest, respectively, as displayed in Fig. 7 with blue lines. The results obtained from the MRMR procedure worked better than those of the other two feature selection strategies, especially for the datasets acquired in April, July, and August, when the performances were almost equivalent to those from VIs_All. The best results of the RReliefF subsets were obtained during the sowing and vegetative stages (April and May), with Ftest being the method that worked the least efficiently. This result might suggest that some vegetation indices described in Table 2 may not contain exploitable information for the GPR model to predict MY. This was observed in particular for the RReliefF algorithm, which, as also observed in the SAT case, was very robust in the choice of VIs. The EVI, GNDVI, NDVI and SR were not selected at all for any of the 12 analysed dates, while the preferred indices were GCVI, NGRDI, TGI and WDRVI. Goh et al. (2022), in accordance with the RReliefF algorithm selection criteria, observed that commonly used vegetation indices, such as NDVI and EVI, were found to be unsatisfactory in predicting the biophysical properties of crops linked to winter wheat growth. However, their results were in disagreement with those of Segarra et al. (2022), who reported a good wheat yield prediction capability when using NDVI or GNDVI indices in an RF. However, the difference between the two supervised ML models could influence such an observation. When considering the definitions of the VIs in Table 2, these results indicated that indices computed using the B8A band (WDRVI and GCVI) were preferred to indices computed with the B08 band (SR, EVI and NDVI), thus suggesting that the part of the NIR spectra that contained the most exploitable information lay between 855 and 875 nm. This observation also agreed with observations in this study from the SAT bands, in which the B8A band was selected a higher number of times that its neighbour B08. It should also be mentioned that some vegetation indices that only contained information from the visible part of the spectrum, such as NGRDI and TGI, were also highly exploited by the feature selection algorithms, unlike the results observed for the SAT scenario. Nevertheless, the results from the VI subsets indicated that, in some cases, the GPR model did not need all 8 VIs as input datasets to achieve its best performance, and that a selection of just four VIs, based on MRMR/RReliefF methods, returned equivalent outcomes.

As in the previous two cases, for the datasets made up of TBIs, the TBIs_All dataset performed markedly better than any subset resulting from feature engineering strategies. This can be observed graphically with the orange lines of Fig. 7. The R^2 values obtained with TBIs_All were on average higher than the scores from TBIs_RReliefF, TBIs_MRMR and TBIs_Ftest, respectively. Such a difference might be linked to the reduction in the number of features (from 36 to 4), which was more marked than in the previous two cases (from 10 to 4 and from 8 to 4 for SAT and VIS, respectively). Again, the RReliefF algorithm showed a remarkable robustness in the choice of its features, and mostly selected TBIs composed of the B07-B06, B11-B05, B12-B06 and B12-B07 bands. These TBIs mostly consisted of combinations of bands in the NIR and in the SWIR ranges, which were the major bands selected by the same algorithm in the SAT dataset. The results agreed with the work presented by Marshall et al. (2022), who observed that the top performing TBIs for MY prediction corresponded to the Red Edge and NIR spectral regions. More specifically, the best performing TBIs for maize included B06 and B11, two of the bands that the RReliefF algorithm also selected several times. The selected TBIs were mostly combinations of bands in different zones of the spectrum, especially SWIR-NIR and SWIR-Red Edge. In addition, TBI B06-B07, which corresponded to a ratio between two bands in the red edge, was selected numerous times. Overall, the performances of the subsets were very similar, and only differed slightly for the April and May dates, where MRMR outperformed the others. The results indicated that the GPR model was able to exploit a TBI feature dataset, thus indicating the validity of this type of approach. In addition, a link between the selection of the best bands from SAT and the bands that composed the most used TBIs was observed. The results also proved that the model performs at its best when all the possible TBI combinations were considered, indicating that any TBI combination was exploitable by the model.

In summary, the use of a complete dataset (SAT_All, TBIs_All, VIs_All) instead of just a subset obtained from feature engineering led to better MY prediction results. However, the RReliefF algorithm showed a remarkable robustness in the selection of its features, and the features it selected had also been identified as important predictors in similar

studies regarding yield estimation with satellite data (Kang et al., 2020; Marshall et al., 2022).

In order to determine the type of spectral data that provided the informative content with the highest predictive potential, a comparison among the complete features datasets (SAT_All, VIs_All and TBIs_All) followed. A clear trend in MY estimation accuracy was observed for all available dates whenever the GPR model was fed with SAT_All, VIs_All and TBIs_All. The results are displayed graphically in Fig. 7, where the R^2 value of each subset defined in Table 3 is plotted over all 12 SAT_i tiles. The best model accuracy for all the analysed dates was observed when the SAT_All dataset was used. In this case, the MY estimation showed R^2 values that ranged between 0.6 and 0.75, with an average of 0.63. The RMSE of the same set of tests was always under 1.7 t ha^{-1}, with a minimum of approximately 1.3 t ha^{-1} and a mean of 1.57 t ha^{-1}. The accuracy obtained when the TBIs_All dataset was used was below that obtained from SAT_All and presented a mean R^2 of 0.54 (20 % lower) and an RMSE ranging from 1.6 to 1.9 t ha^{-1}. The least accurate MY estimations were observed when the GPR model was trained with the VIs_All dataset (mean R^2 of 0.32 and an RMSE ranging from 2.0 to 3.0 t ha^{-1}). These results agreed with the work presented by Hunt et al. (2019), in which Sentinel-2 raw bands were the best type of predictor for wheat yield estimation purposes. Segarra et al. (2022) also observed a better performance when Sentinel-2 bands were used instead of VIs for wheat yield prediction. Segarra et al. (2022) suggested that VIs can suffer from saturation and this might lead to a loss of information, but this does not happen when raw spectral bands are used.

Considering that both the VI and TBI datasets were obtained by performing algebraic operations over the different bands that compose the SAT_All dataset, the above results suggested that the GPR model was better at exploiting raw spectral data when predicting MY. Thus, in order to achieve highly accurate MY estimates from SAT data with a GPR model, it appears appropriate to directly use the normalised raw bands instead of a further feature engineering step. From these observations, it can also be seen that only some VIs might include relevant spectral information that can be correlated with the yield, whereas an extensive analysis, considering all the possible TBIs, resulted in better outcomes.

The MY estimates with the SAT_All dataset returned R^2 values as high as 0.75. These results surpassed the accuracies reported in other MY works. For instance, Marshall et al. (2022) used PLS regression models to estimate MY, considering TBIs and Sentinel-2 bands as predictors, and reported R^2 values as high as 0.6, but generally under 0.5. Desloires et al. (2023) performed an extensive analysis of more than 1000 maize fields in Iowa and Nebraska from 2017 to 2021 to evaluate which spectral information derived from raw Sentinel-2 bands was more able to explain the observed yield variability. They considered several predictor types for this purpose (Sentinel-2 Bands, VIs, biophysical parameters …), which they fed into different ML models, such as SVM, RF and Neural Networks. However, they reported a maximum R^2 value of 0.55. Ji et al. (2022) considered different variables derived from Sentinel-2 (EVI index, Leaf Area index, Land Surface Temperature), and used LASSO, SVM and RF models to estimate MY. They reported R^2 values as high as 0.67, with a mean of 0.63, which was equivalent to the mean obtained in the presented study.

Another interesting trend observed for both the SAT_All and TBIs_All cases was that the accuracy scores showed a certain dependence on the maize life cycle. A constant improvement of MY estimate accuracy was observed during the first months after emergence (May and June), while a peak in performance was seen around mid-June. A decrease in prediction capabilities followed, with the scores reaching lower levels of accuracy than those observed for May. Over July and August, the prediction performance again showed a smooth increase and reached its highest value at the end of August, just before harvesting. However, performances in mid- to late-season (July–August) always returned lower prediction scores than the ones observed in May–June. This trend had already been observed by Hunt et al. (2019) who suggested, as a result of a visual interpretation of the Sentinel-2 images, that the lack of improvement from June to July may be due to the ripening of the crops. This will likely affect the accuracy of yield prediction when Sentinel-2 data are used alone. This coincides with observations from this study. During the vegetative period (May and June), the maize canopy presented a spectral response that was consistent with a vigorous plant status and the GPR model was better able to link it to yield. However, during the reproductive phase (July and August), the canopy size remained constant as the plant gradually senesced while the kernel developed, thereby resulting in a poorer spectral response than that of a healthy plant in the vegetative period. Therefore, during the reproductive phase, it was harder for the GPR model to correlate spectral data with the yield. The work of Kang et al. (2020) supports this hypothesis. They proposed a similar explanation, that is, that surface reflectance has a close correlation with yield during the whole growing season, in particular in the vegetative period. Kang et al. (2020) sustained that such a correlation was especially high for NIR, and they attributed this to the high sensitivity of NIR to the green leaf volume of the canopy, which directly contributes to grain growth. Marshall et al. (2022) also reported better model performances during the vegetative phases of the maize. Cavalaris et al. (2021) observed closer correlations between yield and spectral data when considering flowering at the end of the growing period instead of the whole maize life cycle, again supporting the idea that the vegetative period provides a better spectral response that can be used to link with the yield. This trend implies an in-season estimation of MY around June, i.e. prior to harvesting, could help farmers to guide management decisions and ensure a better harvest.

3.3. Multi-temporal approach/analysis results

The conducted analysis has allowed the most advantageous ML model and input data type to be identified for MY estimation (GPR and SAT_All, respectively). When both were considered, a joint exploitation of spectral data from different maize phenological phases resulted in a clear improvement in the prediction accuracy of MY. As shown in Fig. 8a, results from multi-temporal datasets analysed with the selected data (SAT_1, SAT_2, SAT_3, SAT_{10}, and SAT_{11}) returned R^2 values that ranged from 0.81 to 0.91, which are levels of accuracy that have not previously been reported with mono-temporal approaches. For instance, the multi-temporal dataset that presented the best prediction performance (consisting of combined data from SAT_1, SAT_2, SAT_3, SAT_{10}, and SAT_{11} tiles) returned an R^2 value of 0.91, an increase of 21.3 % compared to the best mono-temporal prediction (R^2 = 0.75). Fig. 8b displays the distribution of the observed vs predicted MY values for this specific case. These results are in agreement with similar works on the topic. Segarra et al. (2022) reported a slight improvement in the estimation of the within-field grain yield when using a whole time-series as input data for their models. Hunt et al. (2019) underlined that the accumulation of Sentinel-2 data over the crop season improved the estimation accuracy of wheat yields, since clear decreases in RMSE were observed when successive Sentinel-2 images were added to their estimation model.

In addition, it can be noted that multi-temporal datasets, consisting of combinations of data from both the vegetative period (SAT_1, SAT_2, SAT_3) and the reproductive period (SAT_{10}; SAT_{12}), in general showed an improved prediction capability, compared to multi-temporal datasets consisting only of time-series of a single phenological phase. Feeding the model with different time series during the vegetative period seems to be a key factor to boost the precision of MY prediction, and, in addition, it allows an in-advanced yield assessment to be made, several months prior to the harvesting phase.

Lastly, it is noted that the sole use of metrics associated with a linear best fit, rather than linear 1:1 fit, is not optimal as it does not account for bias and shift in the model. However, predictions did not tend to show any strong observed shift or bias (Fig. 8b). Future work will consider this

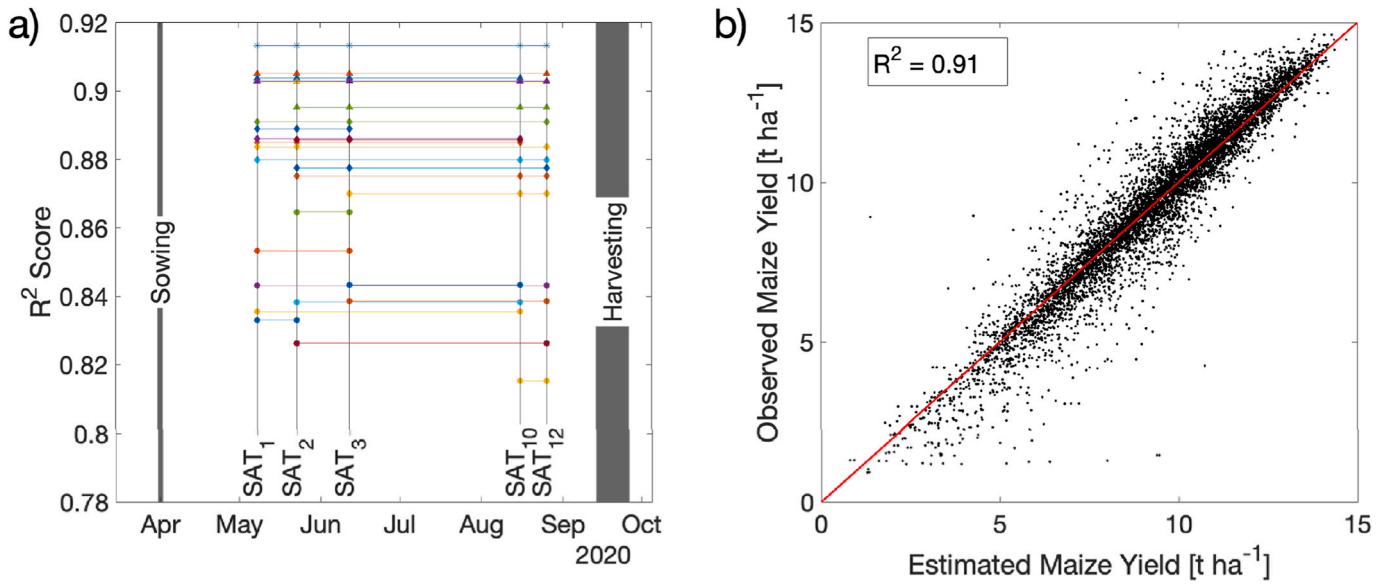

Fig. 8. (a) MY prediction results, in terms of R^2 score, obtained by the GPR model fed with the set of multi-temporal SAT datasets, and (b) scatter plot of observed and predicted MY in the best case, by the GPR model fed with $\{SAT_1, SAT_2, SAT_3, SAT_{10}, SAT_{12}\}$ multi-temporal dataset.

more closely.

4. Conclusions

This work presents an innovative method to estimate maize yields from Sentinel-2 multitemporal-multispectral data by means of feature engineering and supervised machine learning techniques. The study has proved that Sentinel-2 spectral data are able to provide accurate spatial estimates of MY when validated against CANbus in-field data from combine harvesters as ground truth data. An assessment of the ML model showed that GPR was a more suitable ML model, for the prediction task, than SVM, RF or MLR, and the feature engineering study demonstrated that the raw spectral bands from Sentinel-2 were the optimal features to be fed into the models, instead of any derived VI or TBI combination. GPR trained with mono-temporal SAT tiles presented high accuracies (R^2 of up to 0.75), especially during the vegetative period of the crop. However, multi-temporal datasets constructed from single SAT tiles resulted in higher MY estimation accuracies, reaching R^2 values of up to 0.91. These results could have practical applications in precision agriculture, as they allow an in-season MY estimation to be made, prior to harvesting, that could help guide management decisions at the farm scale.

CRediT authorship contribution statement

G. Stefanescu Miralles: Writing – original draft, Validation, Software, Formal analysis, Data curation. **A. Biglia:** Writing – review & editing, Visualization, Formal analysis, Data curation. **D. Ricauda Aimonino:** Writing – review & editing, Validation, Investigation. **M. Mattetti:** Writing – original draft, Project administration, Investigation, Funding acquisition, Data curation, Conceptualization. **P. Gay:** Writing – review & editing, Methodology, Conceptualization. **L. Comba:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Statement on the use of generative AI and AI assisted technologies in the writing process

No generative AI or AI-assisted technologies were used during the preparation of this work.

Funding

This research was partially funded by the project "DATA-BUS - Digital Agriculture Technology to Achieve data to Build User-friendly Sustainability indicators" (PRIN 2020; Prot. 2020SCNF4L) and was partially carried out within the Agritech National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1032 17/06/2022, CN00000022). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abendroth, L. J., Elmore, R. W., Boyer, M. J., & Marlay, S. K. (2011). *Corn growth and development. Ames (IA): Iowa state university cooperative extension service.* PMR 1009.

ArcGIS pro. https://pro.arcgis.com/en/pro-app/3.1/tool-reference/conversion/point-to-raster.htm#GUID-21792F00-9D90-4FBE-9F68-90F98802B802.25/03/2024, (2017).

Arpae. https://aggiornati.arpae.it, (2020)–. (Accessed 30 March 2023).

Atzberger, C., Guérif, M., Baret, F., & Werner, W. (2010). Comparative analysis of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat. *Computers and Electronics in Agriculture, 73*, 165–173. https://doi.org/10.1016/j.compag.2010.05.006

Azar, R., Villa, P., Stroppiana, D., Crema, A., Boschetti, M., & Brivio, P. A. (2016). Assessing in-season crop classification performance using satellite data: A test case in Northern Italy. *European Journal of Remote Sensing, 49*, 361–380. https://doi.org/10.5721/EuJRS20164920

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on computational learning theory* (pp. 144–152). https://doi.org/10.1145/130385.130401

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32. https://doi.org/10.1023/A:1010933404324

Cavalaris, C., Megoudi, S., Maxouri, M., Anatolitis, K., Sifakis, M., Levizou, E., & Kyparissis, A. (2021). Modeling of durum wheat yield based on Sentinel-2 imagery. *Agronomy, 11*, 1486. https://doi.org/10.3390/agronomy11081486

Chauhan, S., Darvishzadeh, R., Lu, Y., Boschetti, M., & Nelson, A. (2020). Understanding wheat lodging using multi-temporal Sentinel-1 and Sentinel-2 data. *Remote Sensing of Environment, 243*, Article 111804. https://doi.org/10.1016/j.rse.2020.111804

Chen, M., Jin, C., Ni, Y., Yang, T., & Zhang, G. (2022). Online field performance evaluation system of a grain combine harvester. *Computers and Electronics in Agriculture, 198*, Article 107047. https://doi.org/10.1016/j.compag.2022.107047
Chung, S. O., Sudduth, K. A., & Drummond, S. T. (2002). Determining yield monitoring system delay time with geostatistical and data segmentation approaches. *Transactions of the ASAE, 45*, 915–926. https://doi.org/10.13031/2013.9938
Copernicus European union©. https://browser.dataspace.copernicus.eu, (2023)–. (Accessed 30 March 2023).
Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other Kernel-based learning methods*. Cambridge University Press. https://doi.org/10.1017/CBO9780511801389
Darra, N., Espejo-Garcia, B., Kasimati, A., Kriezi, O., Psomiadis, E., & Fountas, S. (2023). Can satellites predict yield? Ensemble machine learning and statistical analysis of Sentinel-2 imagery for processing tomato yield prediction. *Sensors, 23*, 2586. https://doi.org/10.3390/s23052586
Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Nicola, L., Rabaute, T., Savinaud, M., Udroiu, C., Valero, S., Bégué, A., Dejoux, J.-F., El Harti, A., Ezzahar, J., Kussul, N., Labbassi, K., Lebourgeois, V., & Koetz, B. (2019). Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world. *Remote Sensing of Environment, 221*, 551–568. https://doi.org/10.1016/j.rse.2018.11.007
Desloires, J., Ienco, D., & Botrel, A. (2023). Out-of-year corn yield prediction at field-scale using Sentinel-2 satellite imagery and machine learning methods. *Computers and Electronics in Agriculture, 209*, Article 107807. https://doi.org/10.1016/j.compag.2023.107807
Ding, C., & Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *Journal of Bioinformatics and Computational Biology, 3*, 185–205. https://doi.org/10.1142/s0219720005001004
Engen, M., Sandø, E., Sjølander, B. L. O., Arenberg, S., Gupta, R., & Goodwin, M. (2021). Farm-scale crop yield prediction from multi-temporal data using deep hybrid neural networks. *Agronomy, 11*, 2576. https://doi.org/10.3390/agronomy11122576
European parliament, 2019 www.europarl.europa.eu/thinktank/en/document/EPRS_BRI, (2019)–. (Accessed 14 March 2024).
Everingham, Y., Sexton, J., Skocaj, D., & Inman-Bamber, G. (2016). Accurate prediction of sugarcane yield using a random forest algorithm. *Agronomy for Sustainable Development, 36*, 27. https://doi.org/10.1007/s13593-016-0364-z
Frampton, W. J., Dash, J., Watmough, G., & Milton, E. J. (2013). Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing, 82*, 83–92. https://doi.org/10.1016/j.isprsjprs.2013.04.007
Ghosh, S. S., Dey, S., Bhogapurapu, N., Homayouni, S., Bhattacharya, A., & McNairn, H. (2022). Gaussian process regression model for crop biophysical parameter retrieval from multi-polarized C-band SAR data. *Remote Sensing, 14*, 934. https://doi.org/10.3390/rs14040934
Gitelson, A. A. (2004). Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of Plant Physiology, 161*, 165–173. https://doi.org/10.1078/0176-1617-01176
Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment, 58*, 289–298. https://doi.org/10.1016/S0034-4257(96)00072-7
Gitelson, A. A., Viña, A., Arkebauer, T. J., Rundquist, D. C., Keydan, G., & Leavitt, B. (2003). Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophysical Research Letters, 30*, 1248. https://doi.org/10.1029/2002GL016450
Goh, B.-B., King, P., Whetton, R. L., Sattari, S. Z., & Holden, N. M. (2022). Monitoring winter wheat growth performance at sub-field scale using multitemporal Sentinel-2 imagery. *International Journal of Applied Earth Observation and Geoinformation, 115*, Article 103124. https://doi.org/10.1016/j.jag.2022.103124
Guyon, I., Elisseeff, A., & Kaelbling, L. P. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*, 1157–1182. https://doi.org/10.1162/153244303322753616
Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., & Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment, 81*, 416–426. https://doi.org/10.1016/S0034-4257(02)00018-4
Harsányi, E., Bashir, B., Arshad, S., Ocwa, A., Vad, A., Alsalman, A., Bácskai, I., Rátonyi, T., Hijazi, O., Széles, A., & Mohammed, S. (2023). Data mining and machine learning algorithms for optimizing maize yield forecasting in central Europe. *Agronomy, 13*, 1297. https://doi.org/10.3390/agronomy13051297
Hunt, M. L., Blackburn, G. A., Carrasco, L., Redhead, J. W., & Rowland, C. S. (2019). High resolution wheat yield mapping using Sentinel-2. *Remote Sensing of Environment, 233*, Article 111410. https://doi.org/10.1016/j.rse.2019.111410
Hunt Jr, E. R., Cavigelli, M., Daughtry, C. S. T., Mcmurtrey III, J. E., & Walthall, C. L. (2005). Evaluation of digital photography from model aircraft for remote sensing of crop biomass and nitrogen status. *Precision Agriculture, 6*, 359–378. https://doi.org/10.1007/s11119-005-2324-5
Hunt Jr, E. R., Doraiswamy, P. C., McMurtrey, J. E., Daughtry, C. S. T., Perry, E. M., & Akhmedov, B. (2013). A visible band index for remote sensing leaf chlorophyll content at the canopy scale. *International Journal of Applied Earth Observation and Geoinformation, 21*, 103–112. https://doi.org/10.1016/j.jag.2012.07.020
Ji, Z., Pan, Y., Zhu, X., Zhang, D., & Dai, J. (2022). Prediction of corn yield in the USA corn belt using satellite data and machine learning: From an evapotranspiration perspective. *Agriculture, 12*, 1263. https://doi.org/10.3390/agriculture12081263
Jordan, C. F. (1969). Derivation of leaf-area index from quality of light on the forest floor. *Ecology, 50*, 663–666. https://doi.org/10.2307/1936256
Kang, Y., Ozdogan, M., Zhu, X., Ye, Z., Hain, C., & Anderson, M. (2020). Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest. *Environmental Research Letters, 15*, Article 064005. https://doi.org/10.1088/1748-9326/ab7df9
Kanke, Y., Tubaña, B., Dalen, M., & Harrell, D. (2016). Evaluation of red and red-edge reflectance-based vegetation indices for rice biomass and grain yield prediction models in paddy fields. *Precision Agriculture, 17*, 507–530. https://doi.org/10.1007/s11119-016-9433-1
Kaya, Y., & Polat, N. (2023). A linear approach for wheat yield prediction by using different spectral vegetation indices. *International Journal of Electronic Governance, 8*, 52–62. https://doi.org/10.26833/ijeg.1035037
Kohrs, R. A., Lazzara, M. A., Robaidek, J. O., Santek, D. A., & Knuth, S. L. (2014). Global satellite composites — 20 years of evolution. *Atmospheric Research, 135–136*, 8–34. https://doi.org/10.1016/j.atmosres.2013.07.023
Kosmowski, F., Chamberlin, J., Ayalew, H., Sida, T., Abay, K., & Craufurd, P. (2021). How accurate are yield estimates from crop cuts? Evidence from smallholder maize farms in Ethiopia. *Food Policy, 102*, Article 102122. https://doi.org/10.1016/j.foodpol.2021.102122
Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift, 15*, 259–263. https://doi.org/10.1127/0941-2948/2006/0130
Li, C., Chimimba, E. G., Kambombe, O., Brown, L. A., Chibarabada, T. P., Lu, Y., Anghileri, D., Ngongondo, C., Sheffield, J., & Dash, J. (2022). Maize yield estimation in intercropped smallholder fields using satellite data in southern Malawi. *Remote Sensing, 14*, 2458. https://doi.org/10.3390/rs14102458
Lyle, G., Bryan, B. A., & Ostendorf, B. (2014). Post-processing methods to eliminate erroneous grain yield measurements: Review and directions for future development. *Precision Agriculture, 15*, 377–402. https://doi.org/10.1007/s11119-013-9336-3
Marshall, M., Belgiu, M., Boschetti, M., Pepe, M., Stein, A., & Nelson, A. (2022). Field-level crop yield estimation with PRISMA and Sentinel-2. *ISPRS Journal of Photogrammetry and Remote Sensing, 187*, 191–210. https://doi.org/10.1016/j.isprsjprs.2022.03.008
Mathworks matlab®. https://www.mathworks.com/products/matlab.html.20/03/2024, (2024).
Mattetti, M., Maraldi, M., Lenzini, N., Fiorati, S., Sereni, E., & Molari, G. (2021). Outlining the mission profile of agricultural tractors through CAN-BUS data analytics. *Computers and Electronics in Agriculture, 184*, Article 106078. https://doi.org/10.1016/j.compag.2021.106078
Mazzia, V., Comba, L., Khaliq, A., Chiaberge, M., & Gay, P. (2020). UAV and machine learning based refinement of a satellite-driven vegetation index for precision agriculture. *Sensors, 20*, 2530. https://doi.org/10.3390/s20092530
Meier, U. (2001). Growth stages of mono-and dicotyledonous plants. *BBCH Monograph. Federal Biological Research Centre for Agriculture and Forestry*.
Mokhtari, A., Noory, H., & Vazifedoust, M. (2018). Improving crop yield estimation by assimilating LAI and inputting satellite-based surface incoming solar radiation into SWAP model. *Agricultural and Forest Meteorology, 250–251*, 159–170. https://doi.org/10.1016/j.agrformet.2017.12.250
Oguntunde, P., Lischeid, G., & Dietrich, O. (2017). Relationship between rice yield and climate variables in southwest Nigeria using multiple linear regression and support vector machine analysis. *International Journal of Biometeorology, 62*, 459–469. https://doi.org/10.1007/s00484-017-1454-6
Panda, S. S., Ames, D. P., & Panigrahi, S. (2010). Application of vegetation indices for agricultural crop yield prediction using neural network techniques. *Remote Sensing, 2*, 673–696. https://doi.org/10.3390/rs2030673
Qader, S. H., Utazi, C. E., Priyatikanto, R., Najmaddin, P., Hama-Ali, E. O., Khwarahm, N. R., Tatem, A. J., & Dash, J. (2023). Exploring the use of Sentinel-2 datasets and environmental variables to model wheat crop yield in smallholder arid and semi-arid farming systems. *Science of the Total Environment, 869*, Article 161716. https://doi.org/10.1016/j.scitotenv.2023.161716
Rasmussen, C. E. (2004). Gaussian processes in machine learning. *Advanced Lectures on Machine Learning, 3176*, 63–71. https://doi.org/10.1007/978-3-540-28650-9_4
Reitz, P., & Kutzbach, H. D. (1996). Investigations on a particular yield mapping system for combine harvesters. *Computers and Electronics in Agriculture, 14*, 137–150. https://doi.org/10.1016/0168-1699(95)00044-5
Robnik-Šikonja, M., & Kononenko, I. (2003). Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning, 53*, 23–69. https://doi.org/10.1023/A:1025667309714
Rosegrant, M. W., & Cline, S. A. (2003). Global food security: Challenges and policies. *Science, 302*, 1917–1919. https://doi.org/10.1126/science.1092958
Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems in the Great Plains with ERTS. *Remote Sensing Center, Texas (USA)*.
Saruta, K., Hirai, Y., Tanaka, K., Inoue, E., Okayasu, T., & Mitsuoka, M. (2013). Predictive models for yield and protein content of brown rice using support vector machine. *Computers and Electronics in Agriculture, 99*, 93–100. https://doi.org/10.1016/j.compag.2013.09.003
Satir, O., & Berberoglu, S. (2016). Crop yield prediction under soil salinity using satellite derived vegetation indices. *Field Crops Research, 192*, 134–143. https://doi.org/10.1016/j.fcr.2016.04.028
Schuster, J. N. (2016). *Measuring volumetric flow rate of grains through a crop harvester to improve crop yield estimation accuracy. Master thesis.* Iowa State University.
Schwalbert, R. A., Amado, T. J. C., Nieto, L., Varela, S., Corassa, G. M., Horbe, T. A. N., Rice, C. W., Peralta, N. R., & Ciampitti, I. A. (2018). Forecasting maize yield at field scale based on high-resolution satellite imagery. *Biosystems Engineering, 171*, 179–192. https://doi.org/10.1016/j.biosystemseng.2018.04.020
Segarra, J., Araus, J. L., & Kefauver, S. C. (2022). Farming and Earth Observation: Sentinel-2 data to estimate within-field wheat grain yield. *International Journal of*

Applied Earth Observation and Geoinformation, 107, Article 102697. https://doi.org/10.1016/j.jag.2022.102697

Spinoni, J., Naumann, G., & Vogt, J. V. (2017). Pan-European seasonal trends and recent changes of drought frequency and severity. *Global and Planetary Change, 48*, 113–130. https://doi.org/10.1016/j.gloplacha.2016.11.013

Statista. www.statista.com/statistics/876232/food-grain-production-volume-in-italy/.14/03/2024, (2022).

Stone, M. L., Benneweis, R. K., & Van Bergeijk, J. (2008). Evolution of electronics for mobile agricultural equipment. *Transactions of the ASABE, 51*, 385–390. https://doi.org/10.13031/2013.24374

Sue Wing, I., De Cian, E., & Mistry, M. N. (2021). Global vulnerability of crop yields to climate change. *Journal of Environmental Economics and Management, 109*, Article 102462. https://doi.org/10.1016/j.jeem.2021.102462

Sun, Y., Liu, R., Zhang, M., Li, M., Zhang, Z., & Li, H. (2022). Design of feed rate monitoring system and estimation method for yield distribution information on combine harvester. *Computers and Electronics in Agriculture, 201*, Article 107322. https://doi.org/10.1016/j.compag.2022.107322

Sureiman, O., & Mangera, C. (2020). F-test of overall significance in regression analysis simplified. *Journal of the Practice of Cardiovascular Sciences, 6*, 116. https://doi.org/10.4103/jpcs.jpcs_18_20

Twardosz, R., Walanus, A., & Guzik, I. (2021). Warming in Europe: Recent trends in annual and seasonal temperatures. *Pure and Applied Geophysics, 178*, 4021–4032. https://doi.org/10.1007/s00024-021-02860-6

Uno, Y., Prasher, S. O., Lacroix, R., Goel, P. K., Karimi, Y., Viau, A., & Patel, R. M. (2005). Artificial neural networks to predict corn yield from compact airbone spectrographic imager data. *Computers and Electronics in Agriculture, 47*, 149–161. https://doi.org/10.1016/j.compag.2004.11.014

Vicente-Serrano, S. M., Lopez-Moreno, J. I., Beguería, S., Lorenzo-Lacruz, J., Sanchez-Lorenzo, A., García-Ruiz, J. M., Azorin-Molina, C., Morán-Tejeda, E., Revuelto, J., Trigo, R., Coelho, F., & Espejo, F. (2014). Evidence of increasing drought severity caused by temperature rise in southern Europe. *Environmental Research Letters, 9*, Article 044001. https://doi.org/10.1088/1748-9326/9/4/044001

Wen, K., Wu, X., & Leatham, D. J. (2021). Spatially smoothed kernel densities with application to crop yield distributions. *Journal of Agricultural, Biological, and Environmental Statistics, 26*, 349–366. https://doi.org/10.1007/s13253-021-00442-6

Yang, C., Everitt, J. H., & Bradford, J. M. (2002). Optimum time lag determination for yield monitoring with remotely sensed imagery. *Transactions of the ASAE, 45*, 1737–1745. https://doi.org/10.13031/2013.11420

Zhang, Y., Yin, Y., Meng, Z., Chen, D., Qin, W., Wang, Q., & Dai, D. (2022). Development and testing of a grain combine harvester throughput monitoring system. *Computers and Electronics in Agriculture, 200*, Article 107253. https://doi.org/10.1016/j.compag.2022.107253

Zhou, X., Kono, Y., Win, A., Matsui, T., & Tanaka, T. S. T. (2020). Predicting within-field variability in grain yield and protein content of winter wheat using UAV-Based multispectral imagery and machine learning approaches. *Plant Production Science, 24*, 137–151. https://doi.org/10.1080/1343943X.2020.1819165