Out-of-year corn yield prediction at field-scale using Sentinel-2 satellite imagery and machine learning methods

Johann Desloires, Dino Ienco, Antoine Botrel

Computers and Electronics in Agriculture 209 (2023) 107807

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Out-of-year corn yield prediction at field-scale using Sentinel-2 satellite imagery and machine learning methods

Johann Desloires [a,b,*], Dino Ienco [b], Antoine Botrel [a]

[a] Syngenta Seeds, Saint-Sauveur, France
[b] INRAE, UMR TETIS, Montpellier, France

ARTICLE INFO

Keywords:
Yield prediction
Machine learning
Sentinel-2
Biophysical parameters
Thermal time

ABSTRACT

Crop yield prediction for an ongoing season is crucial for food security interventions and commodity markets for decisions such as inventory management, understanding yield trends and variability. This work considers corn yield prediction at field-scale with input variables derived from satellite and environmental data. Crop yield data were obtained consecutively from 2017 to 2021 for a total of 1164 fields in the US states of Iowa and Nebraska. We forecast yield "out-of-year", i.e. we test year from using machine learning methods trained on data from other years. This study includes evaluating what spectral information derived from the raw Sentinel-2 bands best explains the observed variability in yields, but also how time is considered for temporal resampling. We found that resampling the annual time series on thermal time and using biophysical parameters estimates increased the R^2 on average by 0.25 to 0.42 when extrapolate is performed on a different year from the ones covered by training samples, compared to using calendar time and other information derived from the Sentinel-2 spectrum.

1. Introduction

Zea mays L. (maize or corn) is one of the most important annual cereal crops in the world's agricultural trade. It is used as food for humans and animals in many parts of the world. It has also recently gained importance as a raw material, especially for the production of biofuels, chemical compounds and pseudo-plastics (García-Lara and Serna-Saldivar, 2019). To meet the growth in global demand without substantial expansion of cultivated land, it is important to continue to increase yields in a sustainable manner to achieve food security, in a way that prevents ecosystem deterioration and fertility loss.

Corn grain yield depends on the genetic potential of the genotype used, soil characteristics, field management practices and agro-climatic factors (Ngoune Tandzi and Mutengwa, 2020; Deines et al., 2021). Estimating actual yields can facilitate management interventions (Hoffman et al., 2015), anticipate supply to meet market demand (Carletto et al., 2015), but also contribute to a broader understanding of the interaction between the environment and the management practices (Beres et al., 2020). The two main approaches that allow government agencies and private industry to estimate pre-harvest yields today are crop growth models and data-driven based models (i.e. joint exploitation of machine learning algorithms and remote sensing data) (Rembold et al., 2013; Ngoune Tandzi and Mutengwa, 2020). Although the former can achieve good results without ground data, it requires a large number of inputs such as agro-meteorological information and farming practices, as well as extensive calibration over the study area under consideration. With limited information, most crop growth models achieve poorly prediction related to crop yield (Palosuo et al., 2011). Conversely, advanced nonparametric methods, coming from the field of machine learning, aim to find a relationship between reference crop yield information (ground data) and remote sensing data without prior assumptions about the underlying data distribution (Behmann et al., 2015). They have been successfully tested to estimate corn yield (Kuwata and Shibasaki, 2016; Mishra et al., 2016; Deines et al., 2021) and they can provide similar results if sufficient ground truth data for training the algorithms is accessible (Deines et al., 2021).

A great deal of attention on multi-year yield prediction is present in the literature, in order to study the generalization of the machine learning model on more diverse environmental conditions (Kamir et al., 2020; Gbodjo et al., 2021; Marszalek et al., 2022). In Deines et al. (2021), the data requirements are estimated to be about 1000 samples to reconstruct historical yields for the years and regions that are also present in their training samples.

In this case, the objective may be similar to interpolating yields for a given season and region, and advanced non-parametric methods are a priori well suited to perform this type of task. Yield predictions can then be used to analyse differences in performance between fields

* Corresponding author at: Syngenta Seeds, Saint-Sauveur, France.
E-mail addresses: johann.desloires@syngenta.com (J. Desloires), dino.ienco@inrae.fr (D. Ienco), antoine.botrel@syngenta.com (A. Botrel).

https://doi.org/10.1016/j.compag.2023.107807
Received 16 December 2022; Received in revised form 3 March 2023; Accepted 24 March 2023
Available online 3 April 2023

in a study area where only limited reference/ground truth data are available. However, a more realistic use case, important for food security interventions and commodity markets, is to forecast yield for an ongoing season for which we do not yet have harvest data, i.e., when extrapolating yield to test samples from years outside of the training samples. According to our literature review, no research has been yet performed a comprehensive comparison of the model's ability for the "out-of-year" crop yields extrapolation task at the field-level, where the objective is to predict crop yields for a target year using other seasons as source data. Machine learning models for crop yield prediction can perform poorly when applied to years and regions outside of their training sample (Lobell, 2013). This may be due to a discrepancy between the distributions of the source and target data, caused in part by different temperature regimes that may accelerate or delay plant growth (Duveiller et al., 2013). The possibility of accounting for seasonal changes in phenology, associated with differences in temperature regime, is also explored using thermal time instead of calendar time.

Satellite-derived spectral information improves the understanding of the contribution of vegetation properties to the yield obtained. They allow reliable spatial and temporal analyses of photosynthetic activity and canopy structural variations. The first attempts to apply satellite imagery for yield prediction date back to the 1970s (Hammond, 1975; Morain and Williams, 1975). These efforts have been massively deployed using satellites at medium/low spatial resolution (MODIS) at the county-to-global scales (Whitcraft et al., 2015; You et al., 2017; Bojanowski et al., 2022), and more rarely at the field-level with higher spatial resolution sensors (i.e. Landsat -8 at 30 m of spatial resolution) (Kang and Özdoğan, 2019; Deines et al., 2021), given its lower revisit time (16-day) and the difficulty of acquiring this type of ground-based data. Since 2016, the Sentinel-2 constellation has provided a promising source of imagery, with 13 spectral bands in various frequencies at 10–60 m of spatial resolution, an image every 5 days and a consistent spatial coverage. Recently, Marszalek et al. (2022) evaluated the use of Sentinel-2 data to perform field-level crop yield prediction and concluded that the use of all raw bands explained the observed variability in winter wheat yields well compared to the use of vegetation indices. Therefore, it requires further study, although it is known that vegetation indices can be easily saturated by a dense canopy (Wang et al., 2012). In addition to these two sources of information, biophysical parameters can also be derived from remote sensing images and have the benefit of mitigating the saturation problem. One solution to obtain their estimation with global coverage is to use the SNAP (Sentinel Application Platform) biophysical processor (Weiss and Baret, 2016). This tool uses an artificial neural network (ANN) pre-trained model, based on the PROSAIL (Jacquemoud et al., 2009) simulation system, to inverse the biophysical variable from optical remote sensing data. The ANN weights are publicly available and accessible. This approach has become the standard one and was the first entry point for estimate biophysical variables from Sentinel-2 data (Hauser et al., 2021).

Commonly used machine learning algorithms for crop yield prediction include empirical regression methods (e.g. Ridge), ensemble learning approaches (e.g. Random Forest, Gradient boosting), Support Vector Regression, and Artificial Neural Networks (Mateo-Sanchis et al., 2019; Kamir et al., 2020; Gbodjo et al., 2021). However, despite the interest in the community, the studies report diverse conclusions with respect to the performances of nonparametric approaches employed in the analysis. It is therefore important to perform comparative studies in order to have a clear understanding of the model behaviours (strengths and weaknesses), or even consolidate the predictions of individual base learners by combining them together (Kamir et al., 2020).

The objective of this study is to further contribute to the understanding of how remote sensing data can be exploited to estimate late-season corn yield using machine learning methods. The relationships between Sentinel-2-derived spectral information, temperature data, and field-level yield data are, here, investigated. More precisely, we propose to (1) explore the possibility of accounting for seasonal changes in phenology by using thermal time instead of calendar time (2) identify the most relevant Sentinel-2-derived spectral information (3) evaluate different complementary machine learning algorithms for such task.

Here, we exploit a multi-year corn production dataset with over 1150 observations of corn field-level yield averages, spanning the U.S. states of Iowa and Nebraska between 2017 and 2021, with an average of 230 fields per year. Experiments are being conducted to evaluate machine learning methods behaviour on "out-of-year" samples, using Sentinel-2 satellite image time series and temperature data derived from ERA-5 climate reanalysis, to monitor crop physiology and temperature regime respectively.

The rest of this article is organized as follows. Section 2 introduces the data available on the study site, the proposed framework associated with experimental settings. The results are reported and discussed in Sections 3 and 4 respectively. Finally, Section 5 concludes.

2. Materials and methods

2.1. Study area

The study site is located in the major corn-producing states of Iowa and eastern Nebraska, located in the Corn Belt of the United States (Fig. 1). They are the first and third largest corn-producing states, with about 16% and 11% in term of total U.S. corn production, respectively. Environmental conditions in this region are characterized by deep and fertile soils, rich in organic matter and nitrogen, with relatively flat terrain. The climate of this region has warm nights, hot days, and well-distributed precipitation during the growing season, which are ideal conditions for growing corn.

2.2. In situ data

2.2.1. Specificities of production fields

Corn production fields were chosen and managed by the Syngenta Production and Supply department[1] for five consecutive years between 2017 to 2021. These fields produced hybrid corn seeds, the major characteristic of which is the crossing of a male-sterile female inbred line, incapable of self-pollination, with a male pollen-donating inbred line. A mechanical detasseling is involved during the season to remove the tassel of female lines and the dates of flower emergence and activity must overlap to ensure proper kernel set (Palagyi and Nemeth, 1996).

2.2.2. Fields description

The study was conducted on 1 162 corn fields, of which 783 fields were located in Nebraska and 379 in Iowa. The size of the selected fields ranged from 3.9 to 276 acres, with a total of 59 323 and 26 741 acres of corn fields grown in these states respectively. 528 fields were irrigated and the percentage is relatively constant from year to year (Table 1). Female corn parental lines were on average planted on day of the year 137 (± 9.3) and harvested on day 263 (± 9.8). These dates were relatively consistent across years, except in 2018 when the average planting and harvest date was on average shortened by one week compared to other years (Table 1). The average length of the season in number of days was 126 (± 8.9), but with large variations between years (Fig. 2(a)).

Corn grain yield was expressed in green bushels per female acre and normalized to a moisture content of 15.5%. Green bushels per female acre (GB/FA) is scaled between 0 and 1 (min–max normalization) for data confidentiality purposes, with an average of 0.53 (± 0.16). The

[1] Syngenta is an international leading science-based agtech company https://www.syngenta.com/en/company.

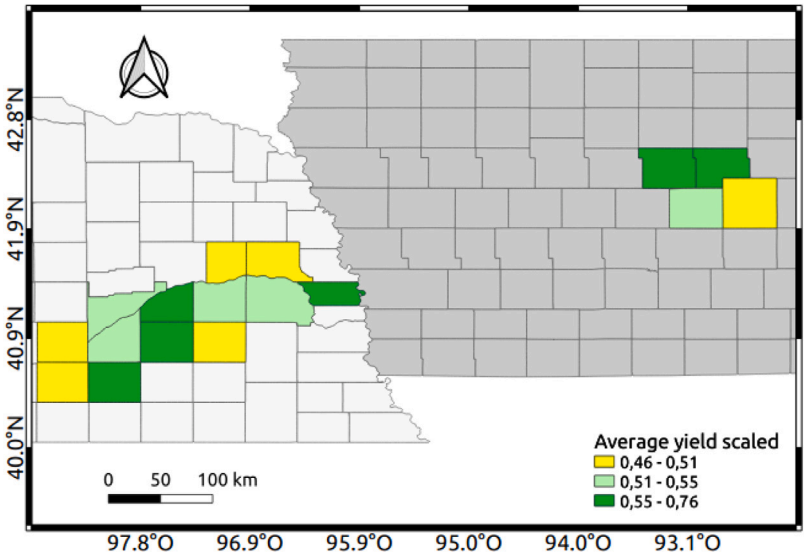
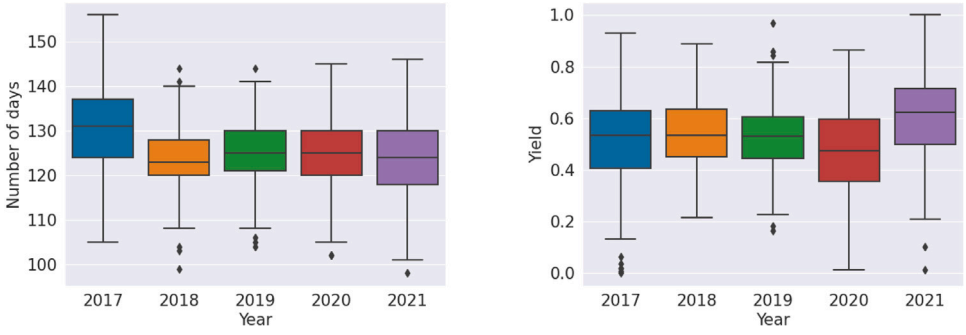


Fig. 1. Seed production fields are spatially distributed over 21 counties in Iowa and eastern Nebraska.



(a) Distribution of growing season length in days by year

(b) Yield distribution by year

Fig. 2. The temporal distributions of season length (b) and yield (c) indicate strong annual effects.

latter showed variability in the temporal distribution of yield production within and between years (Fig. 2(b)) but also spatially between counties (Fig. 1).

Another important variable is the relative maturity of the female parental lines, which was used as an input for the model. The latter measures the time needed to achieve silking and is expressed in number of days (80 to 120 with a step of 5). In principle, a full-season variety can take better advantage of the available heat units and thus give a better yield. However, to vary planting and/or harvesting dates, delay flowering and pollination to protect against drought, earlier maturities are sometimes preferred (Di Salvo et al., 2021).

To summarize, the in-situ features that were used in the yield modelling approach are irrigation, relative maturity of female parental lines and geographic coordinates. The latter, expressed in decimal degrees and rounded by a factor of 1. They were used to characterize agricultural practices that can change from location to location.

Table 1
Summary statistics in-situ data.

Year	# Fields	% Irrigated	Planting date	Harvest date
2017	280	69.6%	137.5 ± 9.3	268.3 ± 12.1
2018	207	74.4%	133.8 ± 7.6	257.3 ± 7.4
2019	187	62.5%	138.9 ± 7.6	264.2 ± 6.7
2020	257	71.2%	140.4 ± 10.2	265.6 ± 8.6
2021	233	71.1%	138.2 ± 9.7	262.4 ± 8.5
Summary	1164	69.3%	137.8 ± 9.3	263.9 ± 9.8

2.3. Sentinel-2 imagery

2.3.1. Acquisition

The Sentinel-2 mission consists of two identical optical satellites, Sentinel-2A and Sentinel-2B, launched respectively on 23 June 2015 and 7 March 2017, that circled the earth with 180° phase. It has a combined repeat cycle of at least 5 days and provides multi-spectral data with 13 bands with a spatial resolution of 10, 20 or 60 m per pixel depending on the band. Bands B2, B3, B4 and B8 have a spatial resolution of 10 m per pixel and cover the visible and near infrared frequencies, while bands B5, B6 and B7, and bands B11 and B12, have a spatial resolution of 20 m per pixel and cover the red-edge and shortwave infrared (SWIR) frequencies respectively. Bands B1, B9 and B10, with a spatial resolution of 60 m per pixel, contribute to atmospheric corrections and geophysical parameters estimation (Bertini et al., 2012) and are therefore not retained in the analysis. Band B2 was not included either, as this channel is affected by atmospheric scattering (Marshall et al., 2022). All the raw data are obtained via the EO-oriented open source Machine Learning (ML) framework eo-learn (Sinergise, 2019). This framework provides a Python client to download and preprocess atmospherically corrected Sentinel-2 data at level 2A from the SentinelHub cloud platform. All raw bands pixels were resampled to 10 m bands reference grid using nearest neighbour interpolation. For each image, the pixels identified as saturated, shaded, or cloudy using the scene classification map (SLC) obtained by sen2cor (Main-Knorn et al., 2017) were discarded in further analyses. Next, we calculate the

percentage of pixels that were previously identified as invalid relatively to the total number of pixels in the field bounding box. Only acquisition dates with more than 50% valid pixels are retained. In addition, since July is the most critical month for corn crop development in the United States (Westcott and Jewison, 2014), fields with less than two acquisition dates between July and August were not included in our study.

2.3.2. Vegetation indices and biophysical parameters

Two-band vegetation indices are widely used in the yield prediction literature. Such indices permit to alleviate possible noise contained in individual bands due to the variation of the soil background, solar geometries and terrain, although they lose some spectral information and can be easily saturated by a dense canopy (Wang et al., 2012). Three relevant vegetation indices for corn yield prediction are calculated (Wang et al., 2014; Schwalbert et al., 2018; Kayad et al., 2019), namely Green Normalized Difference Vegetation Index (GNDVI), Normalized Difference red edge (NDRE) and Normalized Difference Water Index (NDWI). They have the form of NDVI computation, but with the red band replaced by the green for chlorophyll concentration (GNDVI), red-edge for nitrogen status (NDRE) and shortwave-infrared for plant water stress (NDWI), making them complementary.

Biophysical parameters have the benefit of mitigating the saturation problem (Pasqualotto et al., 2019). Using the ANN model pre-trained on the raw Sentinel-2 bands from the SNAP biophysical processor, we have estimated Leaf Area Index (LAI) and Leaf Chlorophyll Content (C_{ab}) While many studies have primarily focused on the assimilation of LAI into crop growth models for corn yield prediction (Lobell et al., 2015; Ines et al., 2013; Novelli et al., 2019) or in empirical approaches (Lykhovyd, 2020), C_{ab} could also be used to better capture the trade-offs of a plant's investment in leaf structure, leaf area, and photosynthetic capacity per area (Hauser et al., 2021). For example, a simultaneous seasonal decrease in C_{ab} and LAI can help assess plant stress due to low nitrogen uptake, a factor limiting plant growth and achievement of genetic yield potential (Lambert et al., 2014). C_{ab} will describe here a limited synthesis or destruction of chloroplasts while LAI indicates a decrease in leaf production coupled with an increase in senescence (Baret et al., 2007).

Because leaf area index in corn fields depends on planting density and plant variety, and hybrids with low leaf area can produce higher grain yields (Lambert et al., 2014), the leaf area index time series is exploited in this study to characterize plant growth and phenology (Verger et al., 2016) rather than productivity. The latter is therefore expressed in values relative to its maximum value over the season, and denoted as LAI_{scaled}.

2.4. Temperature data and thermal time

The developmental rate of corn plants is largely temperature dependent and the corn plant develops best in specific intermediate temperature ranges (Nleya et al., 2019). This is because the many enzymatic activities that occur in the plant are regulated by minimum, maximum, and optimum temperatures (Duveiller et al., 2013). Excessive heat is detrimental to physiological processes such as water stress depletion of soil water, an enhanced root growth at the expense of above-ground biomass, impair flowering and lead to precocious maturity and senescence (Schauberger et al., 2017). Therefore, it is important to consider air temperature in the yield modelling approach, especially for the U.S., as it can significantly affect corn yield during the growing season (Runge, 1968).

The temperature at 2 m above ground (a.g) is associated to each field observation using data from meteoblue API.[2] The data was initially generated from ERA-5 climate dataset, developed by ECMWF,[3] where climate data is simulated in a reanalysis model. It provides hourly estimates of temperature data from 1979 to present in a spatial grid of 31 km.

Thermal time has the advantage of limiting the effects of potential temporal shifts from season to season or location to location (Atwell et al., 1999), which can be detrimental to characterizing actual growing conditions that will then be exploited to predict corn yield (Bojanowski et al., 2022). This will result in smoother and more temporally consistent time series of canopy biophysical variables (Duveiller et al., 2013). Thermal time is typically measured in units called growing-degree days (GDDs), which denotes accumulating mean daily air temperatures at 2 m a.g above a crop-specific threshold. GDDs accumulation over the growing season in days (calendar time) can vary considerably by location and year, whereas the number of GDDs (thermal time) required for corn plants to reach a particular developmental stage is fairly constant (Nleya et al., 2019). Its computation can be written as follows:

$$GDD^{(t)} = \sum_{i=1}^{t} \max\left(\frac{T_{min}^{(i)} + T_{max}^{(i)}}{2} - T_{base}, 0\right) \quad (1)$$

where $T_{min}^{(i)}$ and $T_{max}^{(i)}$ represent minimum and maximum daily temperatures for a given day i while T_{base} stands for base temperature of corn, considered to be 50 °F (10 °C) because corn plants have limited growth below this threshold. The upper threshold temperature is set to 86 °F (30 °C) (Angel et al., 2016). It is assumed that the development rate between these two threshold values is relatively consistent.

GDD is continuously computed for every day during the entire study period running from planting date ($i = 1$) of the corn field to the finishing time ($i = T$) set to 15th of October. Temperature data were grouped into daily averages, minimums, maximums, and number of days with temperatures above 86 °F (30 °C)

2.5. Time series processing

2.5.1. Sentinel-2 data in daily time series

Each Sentinel-2 images time series is represented by the median and standard deviation aggregated at the field-level. To resample time series of satellite images over new time periods to make them comparable in time and space, we applied cubic interpolation to obtain daily time series. The time series were smoothed using the Savitzky-Golay filter, which is typically used to smooth a noisy signal (Savitzky and Golay, 1964) and can better characterize the temporal signals of corn (Shao et al., 2016). The half-width size of the smooth window was set to 15, and a third-degree polynomial was used to set the weighting coefficient.

2.5.2. Temporal characterization of corn growth

The daily time series derived from satellite and thermal data are then aggregated into periods calculated from the sowing date, whether based on calendar or thermal time. For the first, we empirically set periods of 10 days, which gives approximately at most 15 periods. For the second, in order to have approximately the same number of periods as the calendar time, the accumulated GDD values are aggregated with a step of 216 in Fahrenheit (120 in Celcius). For both, we remove the periods where we have more than 90% missing values, giving us a total of 14 periods to cover the season. For the fields where the last period(s) have missing values, we impute them against the observed value of the previous period.

From the calculated periods, we characterize the growth of the plant during the season by using variables derived from Sentinel-2 that allow us to monitor the biomass, in particular by using the LAI estimate. The date of maximum LAI corresponds approximately to the beginning of the flowering–pollination stage (Stone et al., 2000), and we are particularly interested in the values from the time series around this period calculated as follows:

$$t_{max} = argmax\{LAI_{t_i} | t_1 \leq t_i \leq t_n\} \quad (2)$$

[2] https://docs.meteoblue.com/en/weather-apis/dataset-api/dataset-api.

[3] European Centre for Medium-Range Weather Forecasts.

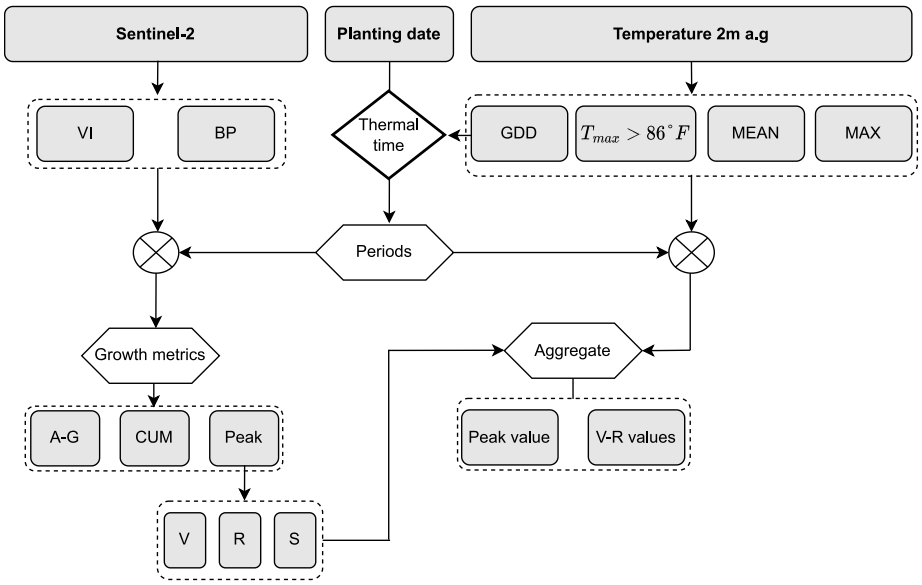


Fig. 3. Flowchart of the methodology for temporal characterization of time series data (feature engineering). For vegetation indices (VI) and biophysical parameters (BP), we compute growth metrics by calculating A-G curve parameters, cumulative values and season peak. The latter also allows us to define the 3 main phases of growth: vegetative (V), reproductive (R) and senescence (S). These phases are used to aggregate the features derived from the temperature.

where t_{max} was the period corresponding to the maximum LAI phenological metric; t_i was the period obtained using calendar or thermal time ; LAI_{t_i} was the LAI value for period t_i; and $argmax f(x)$ was a function that means return x that maximizes $f(x)$. Empirically, we defined the vegetative, reproductive, and senescence plant growth phases as the periods before this peak ($t_1, \ldots, t_{max}$), a period after ($t_{max}, t_{max} + 1$), and the remaining periods ($t_{max} + 1, \ldots, t_n$) respectively. The values associated with this t_{max} period were used to characterize the plant vigour associated with the full leaf development stage, and the number of accumulated GDDs required to reach it.

Because the reproductive stage of corn development, followed by the vegetative stage, are the most sensitive to elevated temperatures for corn yield determination (Hatfield et al., 2011), temperature values were aggregated with respect to peak vegetation, i.e., when LAI is maximum, because earlier maturity is associated with earlier flowering. Therefore, we considered empirically the variables derived from temperature, taking only the aggregated values of the periods before the senescence phase, i.e. ($t_{max} - k$, $t_{max} + 1$), where we define k periods before the peak of the season which has been set here equal to 4.

In addition, we are interested in calculating growth metrics that can summarize crop growth across periods spanning the season. An Asymmetric Gaussian (A-G) curve fitting (Jonsson and Eklundh, 2002), resulting in two functions for the rising (green-up) and falling (maturity) curves, is fitted to the aggregated mean time series of vegetation indices and biophysical parameters. The underlying objective is to extract seasonal growth metrics from its estimated parameters.

$$g(t; a_1, a_2, a_3, a_4, a_5) = \begin{cases} \exp[-(\frac{t-a_1}{a_2})^{a_3}] & if\, t > a_1 \\ \exp[-(\frac{a_1-t}{a_4})^{a_5}] & if\, t < a_1 \end{cases} \quad (3)$$

where, $g(t; a_1, a_2 \ldots . a_5)$ is the Gaussian-type fit function, a_1 is related to position (in time) of the minimum or maximum, a_2 and a_3 are related to the width and flatness of the right half function, and a_4 and a_5 are related to the width and flatness of the left half function. What can have the most impact on yield is stress around flowering, which can then be characterized by accelerated senescence. We therefore combined only the parameters a_2 and a_3 as input variables to the model.

Finally, cumulative values of daily time series were also used to monitor temporal deviations in plant development through the season (Kancheva et al., 2007) and the biomass accumulated over the growing season of the crop, the latter assuming it is proportional to the yield (Tucker et al., 1980).

A flowchart summarizing the approach to characterize the growing season in depicted in Fig. 3.

2.6. Modelling methodology

2.6.1. Yield estimation

The different inputs to this study, detailed in Table 3, are data sources of different nature (temporal and static). Therefore, the simplicity of modelling the data by considering them in tabular form (and not sequentially for the time series) was kept as a central point in this study. Over the past decades, many non-parametric classification and regression algorithms have been proposed, and the choice of their application depends on the following attributes: robustness to training data, sensitivity to changes, data fitting, stability, ability to handle large data sizes, sensitivity to noise, time invested in parameter tuning, and accuracy (Boateng et al., 2020). The performance of the different methods depends strongly on the general characteristics of the data, and a trial-and-error approach is generally used to decide which one to choose.

The superiority of non-linear models over linear models is not guaranteed, and depends on the specific problem being addressed. While non-linear models are typically more suitable for interpolation tasks, linear models may outperform them in extrapolation scenarios (Hastie et al., 2009), and the use of overly complex models may result in overfitting issues, which can negatively impact model performance (Bishop and Nasrabadi, 2006). We therefore introduce Ridge Regression (RR) (Hoerl and Kennard, 1970) as competitor, which is the most popular multiple linear regression method. A penalty term from the L2 norm regularization prevents multicollinearity by shrinking the predictor coefficients.

Most research studies recommend Random Forest (RF) (Breiman, 2001) and Support Vector Regression (SVR) (Cortes and Vapnik, 1995) for nonlinear methods as they are simpler and easier to parametrize, systematically achieving high accuracy and are often easy and faster to deploy. RF has been particularly successful for corn yield prediction using Sentinel-2 data (Kayad et al., 2019), and SVR has been successfully applied in the field of remote sensing (Mountrakis et al., 2011) and crop yield prediction (Kok et al., 2021). We also evaluate Extreme

Gradient Boosting (XGBoost) (Chen and Guestrin, 2016) as a method, which has gained popularity in recent years due to its ability to handle high-dimensional sparse data. XGBoost is an ensemble of decision trees; it consists of a sequential build-off from various decision trees where each tree works to enhance the performance of the prior tree. It is considered superior to previous boosting algorithms because it uses a more regularized model formulation, which allows for better control of overfitting and ultimately improved performance (Cisty and Soldanova, 2018). Huber et al. (2022) have recently shown that a feature-based representation derived from remote sensing can outperform even the most recent deep learning-based yield predictors.

Artificial neural networks (ANNs) have been considered, even when dealing with small data sets, because they have the ability to capture complex relationships between input and output variables and generalize to new data. They can be improved by incorporating regularization techniques, such as weight decay and early stopping. However, the random uncertainty caused by the random initialization of deep learning methods can cause arbitrary errors on test samples that are not part of the training data distribution. To reduce the variability due to initialization, an unweighted average of the outputs of several individual ANNs, randomly initialized and trained on the same dataset, can alleviate this effect or even improve the results of the individual models (Lang et al., 2022). Let M the number of MLPs, the predictions of the ANNs ensemble, denoted as MLP_{ENS}, are calculated as follows :

$$\hat{y}_{MLP_{ENS}} = \frac{1}{M}\sum_{i=1}^{M} = \hat{y}_{MLP_i} \tag{4}$$

where $\hat{y}_{MLP_i}$ are the predicted values from a given individual MLP model, and $\hat{y}_{MLP_{ENS}}$ is the resulting prediction obtained by averaging M MLP models randomly initialized.

We have also added as a method the averaging of predictions from several machine learning models to improve the overall performance and robustness of a predictive model. Indeed, by combining the predictions of several models, the ensemble can mitigate potential weaknesses of individual models and improve the generalizability of the final prediction. A simple and effective approach for creating an ensemble of multiple regression models is the Stacked Averaging Ensemble (SAE) method, which does not require additional training to determine the weights of the ensemble members. Instead, the SAE aggregates the predictions of individual models by taking the average of their outputs, resulting in improved performance compared to using any individual model alone (Gorissen et al., 2009). The ensemble strategy prediction involving RF, SVR, XGBoost and MLP_{ENS} is implemented as follows:

$$\hat{y}_{STACK_{NL}} = \frac{\hat{y}_{RF} + \hat{y}_{SVR} + \hat{y}_{XGBoost} + \hat{y}_{MLP_{ENS}}}{4} \tag{5}$$

where $\hat{y}_{RF}$, $\hat{y}_{SVR}$, $\hat{y}_{XGBoost}$ and $\hat{y}_{MLP_{ENS}}$ are the predicted values of the nonlinear regression methods (namely, RF, SVR, XGBoost and MLP_{ENS}), and $\hat{y}_{STACK_{NL}}$ is the resulting prediction obtained by stacking the unweighted average of these four predictions.

The algorithms RR, RF and SVR are available in the Python library *scikit-learn* (Pedregosa et al., 2011) ; XGboost in the library *xgboost* (Chen and Guestrin, 2016). To avoid overfitting, the hyperparameters of the algorithm are tuned using a grid search, where the root mean square error (RMSE) is defined as an objective function to be minimized in average over the validation sets. The range of values associated with these hyperparameters is reported in Table 3. MLP was implemented through the *Tensorflow* (Abadi et al., 2016) library. Each hidden layer is followed by a batch normalization and a Rectified Linear Unit (ReLU) activation function. The Mean Squared Error (MSE) loss function is used to optimize the model weights. The hyperparameters we have set are the learning rate, the batch size, the number of epochs, the kernel regularizer, the dropout rate and the number of hidden neurons. The combination of hyperparameters tested for each model is presented in the Table 2.

Table 2
Hyperparameters ranges for the different considered approaches. The best parameters setting, for each method, is chosen via Grid Search based on a training/validation strategy.

Method	Hyperparameters	Range
RR	Alpha	{0.01, 0.1, 0.5, 1, 2, 5, 10, 20, 50}
RF	Maximum depth	{3, 4, 5, 6, 7, 8, None}
	Number of trees	{100, 200, 350, 500}
	Maximum features	{'sqrt', 'log2' , 'auto'}
SVR	Kernel	{'poly', 'rbf', 'linear'}
	C	{0.01, 0.1, 0.5, 1, 5, 10, 30, 50, 100}
	Gamma	{'sqrt', 'log2' , 'auto'}
MLP	Learning rate	$\{10e^{-3}, 10e^{-4}, 10e^{-5}\}$
	Batch size	{4, 8, 16, 32}
	Number of hidden neurons	{16, 32, 64, 128, 256}
	Drop-out rate	{0.1, 0.2, 0.3, 0.5}
	Number of epochs	{100, 200, 300}
XGBoost	Learning rate	$\{10e^{-2}, 5e^{-2}, 2e^{-2}, 10e^{-1}, 2e^{-1}\}$
	Maximum depth	{3, 4, 5, 6, 7, 8}
	Subsample ratio samples	{0.45, 0.6, 0.75, 0.9}
	Subsample ratio columns	{0.45, 0.6, 0.75, 0.9}
	Number of trees	{100, 200, 350, 500}

2.6.2. Validation approach

To compare the performance of different possible spectral information derived from Sentinel-2 for the considered yield prediction task (bands and/or vegetation indices, biophysical parameters) and evaluate the robustness of the machine learning algorithms described in 2.6.1, two different evaluation strategies are compared.

The first is a k-fold cross-validation procedure repeated five times, such as the five test sets do not overlap. This allows to generate different splits of the training and test data, and to have a standard deviation of the considered evaluations. Then, given a train/test split of the original data, 10% of the training is used to validate the model, and we construct a total of 10 non-overlapping validation sets. The average of the performance measures over the validation sets is used to select the algorithm's hyperparameter for this data split. Next, we train the algorithm with adjusted hyperparameters found earlier on the entire training set and compute its performance on the test data. We named such validation approach RDM_{CV}.

The second considers the use case where we wish to predict yield for a new season, and thus avoid the unrealistic case where data from the test year is available for model training. Therefore, the model hyperparameters were selected using the leave-one-group-out method, with the groups divided by year. In our data set with five growing seasons, this means that we consider four years for model learning and one for testing. Among these four training years, one is used for validation and the other three for model training. Fig. 4 provides a visual sketch of the LOYO_{CV} evaluation strategy for a given year n. We repeat the procedure on each year devoted to training the model, i.e. here four times. Then, once the model hyperparameters are selected, we train the model across the four training years and the yield predictions for each test year are compared to the crop yields for that season. We named such approach leave-one-year-out CV (LOYO_{CV}). It tests the capability of the model in a real-world scenario in which we force the model to generalize well on a different year using only data gathered from other, different, years.

According to relevant literature in crop yield estimation from optical satellite image (Klompenburg et al., 2020; Marshall et al., 2022), we train the models using four different scenarios, each of them considering different set of information regarding the use of satellite data. The scenarios are:

- *BANDS* for Sentinel-2 raw bands, which may reveal new insights for yield prediction compared to using only VIs (Skakun et al., 2019). Growth metrics were not computed in this case since the raw time series of bands over the season do not exhibit a

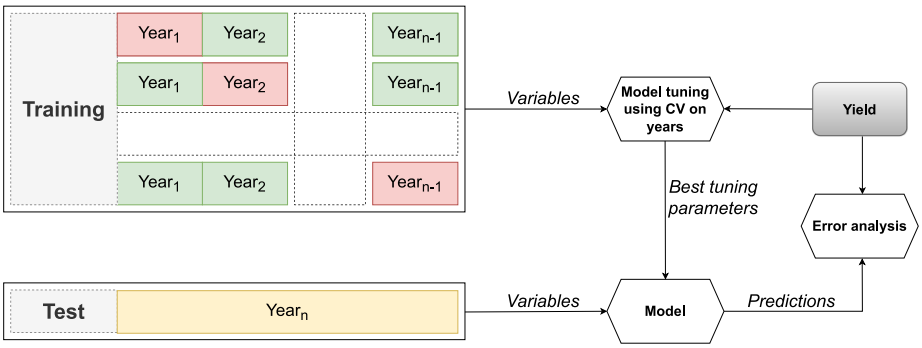


Fig. 4. The Leave-one-year-out cross validation ($LOYO_{CV}$) procedure. First, one year is omitted as a test year (n). In a second step, a grouped cross-validation by year is performed on the years devoted to the training of the model ($1, \ldots, n-1$) to select the hyperparameters, where the green and red colours represent the training and validation years respectively.

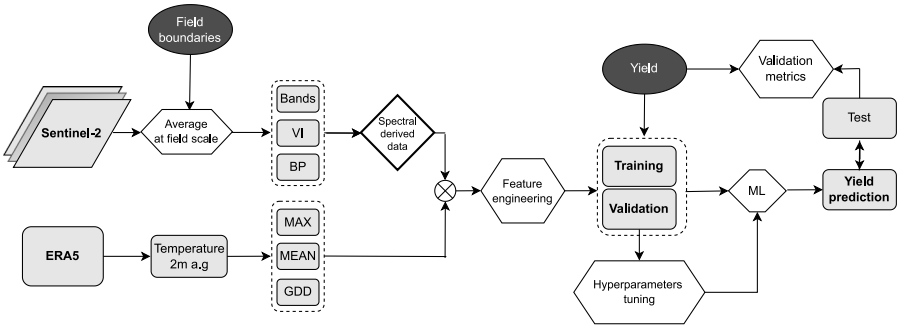


Fig. 5. Pipeline flowchart. Using Sentinel-2, we evaluate spectral information derived from the raw Sentinel-2 bands and combine it with the temperature data. We then evaluate the behaviour of each machine learning method for the performance prediction task using validation metrics on the test samples.

bell-shaped pattern, the latter being a prerequisite for fitting the asymmetric Gaussian parametric function.
- *VI* for the vegetation indices, which is the most widespread approach to determine crop yields. Growth metrics were computed from the GNDVI time series aggregated by period.
- *BANDS+VI* for raw bands and vegetation indices, i.e., combining both *BANDS* and *VI*, to assess whether the two coupled uses can better explain the variability in corn yield.
- *BP* for biophysical parameters estimated using Sentinel-2 raw bands as model input. Growth metrics were computed from the LAI_{scaled} time series aggregated by period.

A summary of the different variables used according to the scenarios is summarized in Table 3. We implement the different scenarios according to the two temporal resampling strategies, which are based on calendar time with 10-day periods and on thermal time described in Section 2.4. Finally, we also study how the two evaluation strategies (RDM_{CV} and $LOYO_{CV}$) affect the performances of the different models for the crop yield estimation task, with the goal to understand how the models behave according to the different scenarios. A flowchart giving a summary of our pipeline is presented in the Fig. 5.

2.6.3. Accuracy assessment

In order to compare model performance between different years, scale-dependent measures such as mean absolute error (MAE) or root mean square error (RMSE) are not well suited since we have different yield distributions between each season, making their respective metrics incomparable to each other (Fig. 2(a)). Instead, we consider relative metrics to compare and assess machine learning model performances. R^2, defined as the square of Pearson's r, estimates the proportion of the variation explained in the observed yield by the predicted yield:

$$R^2 = \left(\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(\hat{Y} - \bar{\hat{Y}})}{\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}\sqrt{\sum_{i=1}^{n}(\hat{Y} - \bar{\hat{Y}})^2}}\right)^2 \quad (6)$$

The Mean Absolute Percentage Error (MAPE) gives an interpretable evaluation of the average accuracy of predictions with respect to average observed yield.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|Y_n - \hat{Y}_n|}{\bar{Y}}100 \quad (7)$$

where Y_n and $\hat{Y}_n$ are the predicted and measured yields respectively, and $\bar{Y}$ is the mean of measured yields.

3. Results

In this section, we present the results of the different data processing, machine learning algorithms and validation strategies described in Section 2. We first examined how the two different validation strategies, namely RDM_{CV} and $LOYO_{CV}$, influence the behaviour of the machine learning methods in Section 3.1. We then compared the use of thermal time versus calendar time for resampling time series observations in Section 3.2. In addition, we assessed the explained yield variation by the models according to different Sentinel-2 derived spectral information, namely *BANDS*, *VI*, *BANDS+VI* and *BP*, previously introduced in Section 3.3. Regarding the machine learning algorithms introduced in Section 2.6.1, we reported their behaviours according to the average performances obtained, as well as their dispersion around the mean, obtained in term of R-squared (R^2) in Section 3.4. Successively, we provided an in-depth investigation of their results in terms of MAPE to verify if the variation of this metric was consistent with the R^2 reported in Section 3.5. Finally, in Section 3.6, we assess the machine learning models to years outside of their training sample and analysed the results for each growing season.

Table 3
Variables used in crop yield estimation during 2017–2021.

Type	Term	Definition	Associated growth stage
Air temperature 2 m a.g	T_{mean}	Daily mean temperature	Vegetative to reproductive
	T_{max}	Daily maximum temperature	
	T_{min}	Daily minimum temperature	
	$Heat_{days}$	Number of days daily $T_{max} > 30$ °C	
Biophysical parameters	C_{ab}	Leaf chlorophyll content (μ_g cm^{-2})	Whole season
	LAI_{scaled}	Leaf area index relative	
	C_{abMAX}	Maximum value C_{ab}	Peak
	C_{abSTD}	Standard deviation C_{ab}	
Vegetation indices	NDWI	$\frac{B8-B11}{B8+B11}$	Whole season
	GNDVI	$\frac{B8-B03}{B8+B03}$	
	NDRE	$\frac{B8-B05}{B8+B05}$	
	$GNDVI_{MAX}$	Maximum value GNDVI	Peak
	$GNDVI_{STD}$	Standard deviation GNDVI	
Growth metrics	a_2	Width right-half function	Senescence
	a_3	Flatness of the right half function	
	TS_{CUM}	Accumulated time series (integral)	Whole season
	GDD_{PEAK}	Accumulated degree days	Peak
In-situ	ReM	Relative maturity	\
	Lon	Longitude	\
	Lat	Latitude	\
	Irr	Irrigation	\

Table 4
Performances in term of R^2.

CV	Resampling	Source	RR	RF	SVR	XGBoost	MLP	MLP_{ENS}	$STACK_{NL}$
RDM_{CV}	Calendar	*BANDS*	0.37 ± 0.04	0.44 ± 0.02	0.45 ± 0.04	0.44 ± 0.02	0.41 ± 0.02	0.44 ± 0.03	0.49 ± 0.04
		VI	0.36 ± 0.03	0.45 ± 0.03	0.43 ± 0.04	0.43 ± 0.01	0.38 ± 0.01	0.41 ± 0.03	0.47 ± 0.03
		BANDS+VI	0.37 ± 0.04	0.44 ± 0.03	0.45 ± 0.04	0.45 ± 0.01	0.42 ± 0.01	0.45 ± 0.00	0.48 ± 0.01
		BP	0.33 ± 0.04	0.47 ± 0.02	0.44 ± 0.03	0.46 ± 0.02	0.43 ± 0.02	0.44 ± 0.02	0.48 ± 0.03
	Thermal	*BANDS*	0.41 ± 0.03	0.45 ± 0.02	0.49 ± 0.02	0.55 ± 0.01	0.49 ± 0.00	0.51 ± 0.01	0.52 ± 0.01
		VI	0.33 ± 0.02	0.48 ± 0.02	0.47 ± 0.03	0.52 ± 0.01	0.46 ± 0.00	0.48 ± 0.00	0.49 ± 0.03
		BANDS+VI	0.34 ± 0.03	0.47 ± 0.02	0.48 ± 0.02	0.54 ± 0.01	0.47 ± 0.00	0.49 ± 0.01	0.50 ± 0.02
		BP	0.36 ± 0.02	0.52 ± 0.02	0.51 ± 0.02	**0.55 ± 0.02**	0.48 ± 0.02	0.51 ± 0.02	0.53 ± 0.01
$LOYO_{CV}$	Calendar	*BANDS*	0.20 ± 0.08	0.20 ± 0.10	0.20 ± 0.06	0.23 ± 0.12	0.19 ± 0.10	0.22 ± 0.10	0.23 ± 0.07
		VI	0.20 ± 0.08	0.21 ± 0.11	0.20 ± 0.06	0.25 ± 0.13	0.22 ± 0.09	0.24 ± 0.10	0.25 ± 0.11
		BANDS+VI	0.21 ± 0.07	0.20 ± 0.11	0.20 ± 0.07	0.28 ± 0.14	0.19 ± 0.09	0.21 ± 0.08	0.26 ± 0.08
		BP	0.18 ± 0.09	0.29 ± 0.11	0.24 ± 0.11	0.31 ± 0.13	0.29 ± 0.07	0.31 ± 0.08	0.30 ± 0.10
	Thermal	*BANDS*	0.22 ± 0.12	0.20 ± 0.10	0.32 ± 0.08	0.22 ± 0.15	0.23 ± 0.19	0.25 ± 0.19	0.28 ± 0.11
		VI	0.26 ± 0.06	0.23 ± 0.11	0.32 ± 0.07	0.28 ± 0.11	0.23 ± 0.12	0.25 ± 0.12	0.31 ± 0.06
		BANDS+VI	0.28 ± 0.08	0.25 ± 0.10	0.33 ± 0.09	0.27 ± 0.15	0.25 ± 0.13	0.27 ± 0.13	0.34 ± 0.09
		BP	0.24 ± 0.06	0.37 ± 0.06	0.38 ± 0.07	0.38 ± 0.09	0.36 ± 0.06	0.38 ± 0.06	**0.42 ± 0.07**

3.1. Comparison of validation strategies: mixing seasons vs. predicting a new season

Before comparing the two validation strategies RDM_{CV} and $LOYO_{CV}$, we briefly recall how they were built. For both, approximately 20% of the dataset was devoted to the test set. The procedure was repeated five times to obtain average results and standard deviations. Therefore, each model presented in Section 2.6.1 had its set of hyperparameters (Table 3) for a given training/test split, which was obtained based on the average of the model scoring on the validation sets. We present the average metrics obtained for each method and their respective standard deviation in term of R^2 and MAPE on the test set in Tables 4 and 5, respectively. We observed that all methods achieved the best performance with RDM_{CV} compared to $LOYO_{CV}$, where the R^2 at the field-level decreased from 0.55 to 0.42 and resulted in a corresponding increase in MAPE (14.0 to 15.1%). This allowed us to quantify the degradation in performance when we forecast corn yields out-of-year at the field-level.

3.2. Comparison of calendar and thermal time temporal resampling

In order to evaluate the resampling strategies between calendar time and thermal time, we recall that we considered the same number of periods for both methods to make the comparisons fair in terms of the number of input variables for the model (Section 2.6.2). Before analyse the predictive capacity, a comparison of yearly NDVI time series profiles between calendar and thermal time resampling was illustrated in Fig. 6. NDVI resampled over calendar time depicted more shifts in crop development along the season, which was well illustrated here by differences in annual temperature patterns. For example, reproductive stage appeared earlier in 2018 and senescence later in 2019, this time shift is reduced considering thermal time.

Concerning the estimation of corn yield, we observed that regardless of the Sentinel-2 data source and validation strategy, the largest accuracy increases, for all the competing approaches considering both validation strategies, happened when the time series data were resampled over thermal time, as highlighted by the experimental results. In particular, it seemed to allow a better generalization over a new season, which explained up to 42% of corn yield variability, contrary to 31%

Table 5
Performances in term of MAPE.

CV	Resampling	Source	RR	RF	SVR	XGBoost	MLP	MLP_{ENS}	$STACK_{NL}$
RDM_{CV}	Calendar	*BANDS*	16.2 ± 0.7	15.4 ± 1.0	15.0 ± 0.9	14.6 ± 0.8	14.9 ± 0.9	14.4 ± 0.8	14.5 ± 0.8
		VI	16.4 ± 0.1	15.4 ± 1.0	15.2 ± 1.0	14.9 ± 0.9	15.1 ± 1.0	14.5 ± 0.7	14.9 ± 1.0
		BANDS+VI	16.4 ± 1.0	15.4 ± 1.0	15.0 ± 0.9	14.6 ± 0.9	14.8 ± 0.3	14.3 ± 0.2	14.7 ± 0.4
		BP	16.8 ± 0.6	14.9 ± 0.7	14.5 ± 0.7	14.7 ± 0.6	14.6 ± 0.6	14.4 ± 0.5	14.5 ± 0.6
	Thermal	*BANDS*	15.7 ± 1.1	15.3 ± 0.9	14.1 ± 0.5	14.9 ± 0.7	15.2 ± 1.1	14.1 ± 0.8	**13.9 ± 0.6**
		VI	17.0 ± 1.4	15.8 ± 001	15.8 ± 0.7	15.3 ± 0.7	15.4 ± 0.3	15.2 ± 0.1	15.3 ± 0.8
		BANDS+VI	17.0 ± 1.3	15.8 ± 1.2	15.6 ± 0.7	15.2 ± 0.5	15.3 ± 0.2	15.1 ± 0.1	15.2 ± 0.3
		BP	16.7 ± 1.0	14.0 ± 0.7	14.3 ± 0.7	14.2 ± 0.8	14.9 ± 0.7	14.7 ± 0.6	14.1 ± 0.7
$LOYO_{CV}$	Calendar	*BANDS*	18.2 ± 3.2	18.2 ± 3.2	18.4 ± 3.2	17.3 ± 3.8	17.9 ± 4.1	17.5 ± 4.0	17.8 ± 3.5
		VI	18.0 ± 3.4	18.0 ± 2.9	18.0 ± 3.4	16.9 ± 2.9	17.1 ± 2.8	16.7 ± 2.7	17.4 ± 2.5
		BANDS+VI	18.2 ± 3.6	18.0 ± 2.9	18.0 ± 3.4	17.5 ± 2.7	17.6 ± 3.0	17.4 ± 3.0	17.7 ± 2.8
		BP	18.2 ± 3.2	16.8 ± 2.3	17.1 ± 2.5	16.6 ± 2.8	16.7 ± 3.0	16.5 ± 3.0	16.4 ± 2.4
	Thermal	*BANDS*	17.7 ± 4.1	18.0 ± 3.3	16.8 ± 3.5	17.0 ± 3.5	17.1 ± 3.6	16.9 ± 3.7	17.1 ± 3.9
		VI	17.6 ± 3.3	17.8 ± 3.1	16.8 ± 3.5	16.9 ± 3.8	17.3 ± 4.4	17.0 ± 4.3	16.5 ± 2.8
		BANDS+VI	17.6 ± 3.8	17.8 ± 3.1	16.8 ± 3.5	16.7 ± 3.7	17.1 ± 3.7	16.7 ± 3.6	16.6 ± 3.4
		BP	17.5 ± 3.2	15.9 ± 2.7	15.7 ± 2.1	15.5 ± 2.4	15.6 ± 2.8	15.4 ± 2.6	**15.2 ± 2.3**

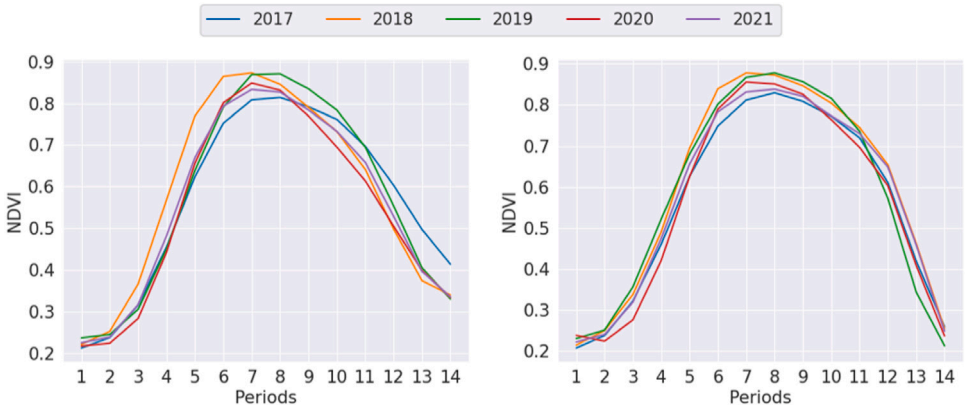


Fig. 6. NDVI time series profile disaggregated by year and resampled over calendar time (left) and thermal time (right). Thermal resampling clearly allows to reduce the time shift of the crop growth between seasons.

using calendar time. This also reduced the relative mean error from 16.5% to 15.2%. This performance gain was also notable under the case of RDM_{CV} evaluation strategy, but to a lesser degree, with 55% of the variability explained against 48%.

3.3. Comparison on sentinel-2 derived spectral information

We then compared the different variables derived from the Sentinel-2 raw bands (Section 2.6.1) as inputs to the machine learning models, namely the raw bands (*BANDS*), vegetation indices (*VI*), combined raw bands and vegetation indices (*BANDS+VI*), and estimated biophysical variables (*BP*). We observed that there was no systematic benefit in terms of model evaluation to add vegetation indices to band time series as model inputs (*BANDS+VI*) rather than considering only bands (*BANDS*). However, we noticed that the right choice of Sentinel-2 derived spectral information depends on the model evaluation strategy, i.e., whether or not seasons were mixed in the test dataset. While we had little difference with RDM_{CV} validation, using *BP* as input performed most of the time better with $LOYO_{CV}$ validation for all nonlinear methods, obtaining an R^2 of 42% against at best 31% for the other input variables.

3.4. Comparison of model algorithms

RR linear model generally achieved lower performance to the non-linear methods, which suggests that considering independent linear additive effects was not sufficient to explain the variance in corn yield in our study case. This was especially true for the variable choice *BP*, where the model was not able to better exploit the relationship with yield for a new season, unlike the non-linear models. Conversely, RF, SVR, XGBoost and MLP obtained very similar results in average, except on the thermal time using the derived variables *BANDS*, *VI* and *BANDS+VI*, where SVR got a higher R^2.

Regarding the neural network approach MLP, while it did not outperform in average the other standard algorithms used for yield prediction, averaged predictions from MLP trained multiple times on the same training set (MLP_{ENS}) slightly improved the average R^2.

Another ensemble learning strategy introduced earlier was to consider the average of the individual nonlinear models RF, SVR, XGBoost and MLP_{ENS}, namely $STACK_{NL}$. Compared to the average performance of the best individual model, we observed with the ensemble learning strategy $STACK_{NL}$ an increase in terms of R^2 from 0.38 to 0.42. This improvement can be justified by the differences between the stochastic learning procedure intrinsic to each machine learning model, whose adjustment mechanism on the data can lead to different errors on the samples. This phenomenon can be even more exacerbated for the samples of the test year, whose distribution of explanatory variables is different from that of the training years. Averaging the predictions allowed us to obtain more robust predicted values by averaging the deviations of each individual model around the true observed value.

3.5. Comparison of variations in model evaluation metrics

We observed that for $LOYO_{CV}$ validation strategy, we generally obtained the same choice in terms of data processing (temporal resampling and Sentinel-2 derived variables), but also for the selection of

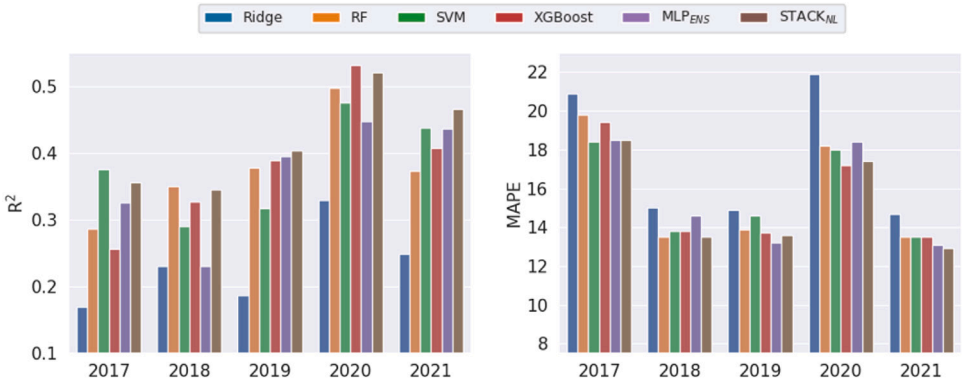


Fig. 7. Comparison of model performance when we seek to predict yield for a new season (temporal cross-validation) using biophysical variables in terms of R-squared (left) and MAPE (right) for RR, RF, SVR, XGBoost, MLP$_{ENS}$ and STACK$_{NL}$ methods.

the machine learning model. A gain in terms of R^2 was also confirmed by a decrease in terms of MAPE, which was highlighted in grey in the Tables 4 and 5. With RDM$_{CV}$ validation strategy, assessing the methods performances under different Sentinel-2 derived variables showed that using estimated biophysical parameters (*BP*), or bands (*BANDS*), resulted in a similar absolute relative error, while the former performed better in term of R^2.

3.6. Fine evaluation of the best model applied to years outside of their training sample

A common scenario was to predict the yield for a year whose yields had not yet been observed, during the model training, to assess for a new season whether we were likely to be under- or over-production. We therefore focused on the results obtained by the best yield estimation method previously obtained with LOYO$_{CV}$ evaluation strategy, namely the STACK$_{NL}$ method, where we used the thermal calendar as a temporal resampling method, and the biophysical variables derived from Sentinel-2 (*BP*) as model inputs. To quantitatively and qualitatively evaluate the predictions obtained over years outside of their training sample, we examined the extent to which the predicted and observed returns lie on the 45° line through the origin, but also whether their respective medians and interquartile ranges were close.

The results per growing season of each method were depicted in Fig. 7. We noticed that, although individual models RF, SVR, XGBoost and MLP$_{ENS}$ were able to describe much of the variation in yields, we did not had a clear winner between them over all seasons. For example, RF was able to explain 35% of the variance in 2018 against only 25% for SVR and MLP$_{ENS}$, while for 2021, the latter two were able to explain 40% of the variance against 35% for the former. This can be explained by the fact that both SVR and MLP require the input data to be scaled, and the latter can be affected by extreme values. Therefore, the strategy of averaging predictions, as an ensemble learning approach with STACK$_{NL}$, was successful in smoothing the performance of individual models. It even outperformed each model for 2020 and 2021. This resulted in an average R^2 of 42%, instead of 38% with the best individual non-linear methods. Fig. 8 depicts the scatter plots of observed corn yields versus predicted yields. The degree to which predicted and observed yields lie on the 45° line through the origin was satisfactory for 2020 and 2021, while for 2017 and 2018, the predicted values showed a positive correlation but struggles to explain much of the variance around the average observed corn yield values.

In addition, we analysed whether the methods were able to capture the yearly effect, i.e. if it is able to predict a season in which the distribution of corn yield values is significantly different from what has been observed on average in other years. Fig. 9 depicts the distribution of observed and predicted performance values by displaying boxplots disentangled by year. Although the interquartile range is much lower for the predicted values than for the observed values, the median values per year are relatively close. This suggests that our approach can provide information to help stakeholders make agricultural forecasts of possible over- or underproduction regarding the current season.

4. Discussion

Predicting corn yield is a notoriously challenging task and there is no standard machine learning practice for predicting corn yield at the field scale using Sentinel-2 data, which is mainly due to the difficulty of acquiring such data. In our study, we developed a field-level corn yield prediction method capable of extrapolating to years different from those observed by coupling freely available information from high-resolution multi-temporal remote sensing imagery and temperature data.

4.1. Comparison with prior studies

4.1.1. Selection of exclusive data

In Section 3.1, we pointed out that the use of the RDM$_{CV}$ evaluation strategy overestimated the ability of the model to generalize corn yield estimations on unseen fields. This may be because our georeferenced corn yield data show spatial autocorrelation for a given year. Therefore, randomly sampling these geospatial data to form training and test sets can inflate the validation parameters of machine learning models, whose goal is to evaluate their ability to generalize on new data (Kattenborn et al., 2022). Conversely, LOYO$_{CV}$ allowed us to consider a more realistic use case in which we want to predict yield based on field data from different years. To the best of our knowledge, few studies have focused on the exclusive selection of data to assess crop yield prediction at the field-level, i.e. when years of the test samples are not in the training samples (extrapolation). The comparison of the results from the two validation strategies allowed us to quantify the degradation of results when attempting to predict for a new season, but also to see if the findings are similar with respect to the choice of input variables and their preprocessing.

In Section 3.2, we have shown that the use of the thermal calendar benefits both validation strategies. Similar conclusions regarding the use of thermal rather than calendar time have been drawn, both at the region (Duveiller et al., 2013) and field (Durgun et al., 2020) scales, and this was also confirmed in our study case for corn at field scale. The gain is even more pronounced when we consider the case of extrapolation of seasons, where the use of thermal resampling reduced the domain mismatch between training and test years, resulted in higher explanatory power of the yield estimate, probably due to a more consistent representation of crop physiology.

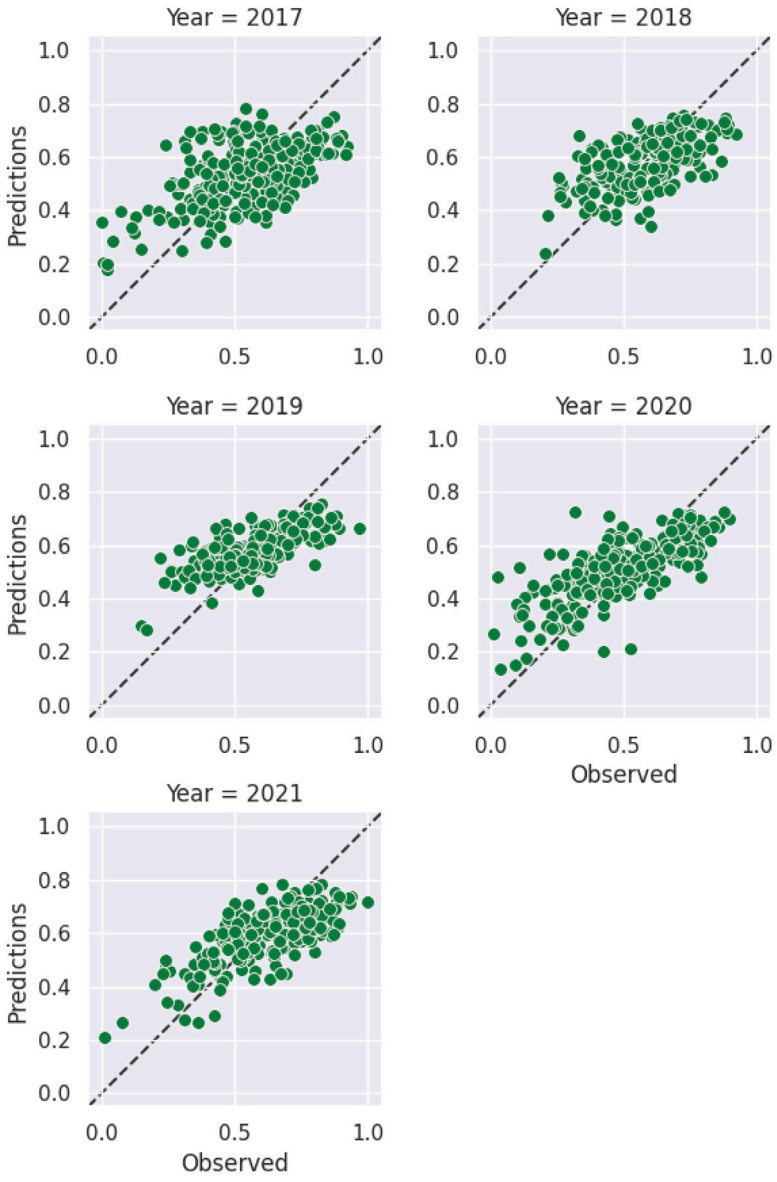


Fig. 8. Scatter plots of observed yield *versus* predicted yield values per growing season with STACK_{NL}.

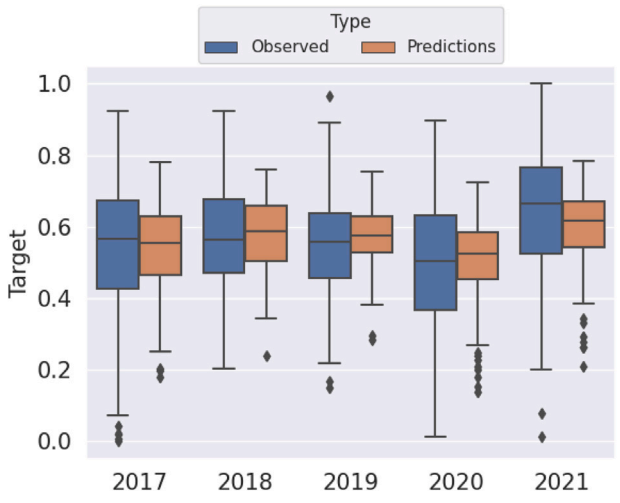


Fig. 9. Distribution of observed yield and predicted yield values with STACK_{NL} method per growing season.

4.1.2. Sentinel-2 spectral derived information

In Section 3.3, the use of estimated biophysical parameters (*BP*) was interesting to increase the signal-to-noise ratio compared to the raw bands when attempting to predict for a new season. They therefore seem to allow the model to better extrapolate the spectral response that characterizes the growth conditions of the plant, unlike vegetation indices that are less sensitive to a plant with high physiological and biochemical content (Wang et al., 2012).

This finding, related to the processing or combination of Sentinel-2 derived spectral information, were recently made for within-field wheat grain yield estimation in Segarra et al. (2022), where the use of LAI estimation gave better results than the use of raw bands and/or vegetation indices. Li et al. (2022) also compared the performance between raw bands and vegetation indices for winter wheat crop yield prediction at field scale and they found that the usage of all raw bands outperformed the index-based approach. For corn yield estimation, Kayad et al. (2019) have compared several vegetation indices and found that GNDVI was the best performing VI, while Li et al. (2022) have rather concluded on the importance of red edge-based VIs. A compromise between these two studies was found in Marshall et al. (2022), where the narrow red bands combined with the wide NIR bands of Sentinel-2 at vegetative

and reproductive stages performed well in estimating corn yield, as a result to its strong linear correlation with canopy chlorophyll content. This was also confirmed in our study through the direct use of Leaf Chlorophyll Content (C_{ab}) estimates, which had the greatest predictive power for the extrapolation of corn yield.

4.1.3. Yield estimation methods

Many mechanistic or empirical approaches have been developed to model the relationships between observable information and corn yield, but there is not yet a consensus on which approach is most appropriate for a given application (Estes et al., 2013). Kang and Özdoğan (2019) combined satellite observations in a modified version of SAFY (Duchemin et al., 2008) crop simulations and explained similar field variability (R^2 = 0.41) without ground-calibrated data. More recently, SCYM (Deines et al., 2021) allowed to further improve such approach for field-level yield means accuracy (R^2 = 0.45) using APSIM (Holzworth et al., 2014) crop simulations. They also compared with RF trained on ground-calibrated data and they found the latter can achieve similar performance than SCYM when trained on at least one thousand ground observations randomly sampled, but performed poorly when tested on years not represented in the training data. We can therefore suggest that our approach trained on calibrated ground data (R^2 = 0.42) achieved very similar performances to SCYM. This is encouraging since we are working with a limited amount of reference field data (less than a thousand) and assess the model performances on out of distribution observations coming from year not covered by the training data.

Nonlinear machine learning algorithms had shown their ability in prior studies to capture complex interactions between spectral information and corn grain yield (Kayad et al., 2019; Deines et al., 2021). RF is often presented as a simple and better performing method for this task (Deines et al., 2021). Van Klompenburg et al. (2020) conducted a systematic review of the literature on crop yield prediction and found that the most widely used algorithm is that of artificial neural networks, specifically the Convolutional Neural Network (CNN), followed by Linear Regression, Support Vector Machines and Gradient Boosting Trees. Although CNN model, which uses a convolution operation to capture the time dependencies, achieve state-of-the-art results (You et al., 2017; Khaki et al., 2020) at the county level, we did not use them in our comparative study at the field-level. The limited amount of reference field data, as well as the complexity of how jointly integrate, in a neural network architecture, both time series as well as static variables (Table 3), explain this choice. In our study, nonlinear models were more effective in understanding the relationship between the crop yield and input data. Looking more closely at the model evaluation metrics when considering the extrapolation of corn yield for each season in Fig. 7, we do not have clear winner between RF, SVR, XGBoost and MLP_{ENS}. Rather than having to choose an individual model, the Stacked Averaging Ensemble ($STACK_{NL}$), with diverse and complementary machine learning approaches, smoothed out or even outperformed the differences in performance of individual models between the seasons.

4.2. Limitations

Reliance on satellite-derived data such as vegetation indices (VIs) or biophysical parameters (e.g. LAI, C_{ab}) along with data-driven approaches only partially capture the complete range of actual crop yield variation. This is because optical data primarily reflects the photosynthetically active biomass within fields, but not necessarily the grain yield which depends on several factors (Rudorff and Batista, 1990). Enriching remotely sensed imagery with additional relevant explanatory variables related to genotype, soil characteristics, field management practices, and agroclimatic factors (Xu et al., 2019) would help reduce the bias associated with omitted variables. Our purely data-driven approaches risk to ignore the underlying physiological mechanisms of plant growth and yield as it relies on only a small part of the yield determinants, namely temperature and optical imaging to monitor plant growth and health. Obtaining additional relevant in-situ data can be expensive and limit the applicability of the model to a specific dataset. In addition, incorporating a large number of agroclimatic factors could also lead to over-fitting and/or to the problem of the curse of dimensionality, which can occur in particular when the number of input variables is greater than the sample size of the dataset (Bishop and Nasrabadi, 2006).

4.3. Future work

The work presented represents a significant advance in understanding the use of Sentinel-2 data for corn yield prediction at field scale. Extending the study with more reference data, with other regions, or even considering other crops, would obviously be beneficial to validate the use of the thermal calendar and the biophysical variables. Concerning the choice of input variables, the most obvious way to improve the results is to include additional agroclimatic factors in our framework. One way would be introducing the output variables of the crop model simulations as inputs, to further exploit environmental factors that may affect corn yield. These could indirectly inject human knowledge on crop yield that has been empirically validated.

From a machine learning perspective, the results could also be improved if we better exploit the temporal dimension of our data, for example by design neural network architecture using CNNs to manage time series data. Finally, it would be interesting to obtain extrapolation of yields over the season rather than at the end. However, this task may be challenging as a stress on the vegetation, such as a water deficit, at a given time of the season is only visible with a certain delay through optical satellite imagery. The difficulty and lack of studies on this topic led us to believe that a comprehensive study of in-season yield prediction using this ground truth at the field-level could be an interesting avenue for future work.

5. Conclusion

This study analysed and compared the performance of data-driven methods for seasonal corn yield estimation from ground-based data in the U.S. using Sentinel-2, temperature and in-situ data. The capacity of estimating yield based on this corn yield seeds production dataset was shown to be influenced by various aspects, namely : (1) the way in which temporal dynamic is considered (thermal or calendar); (2) the Sentinel-2 derived spectral information (bands and/or vegetation indices, biophysical parameters); (3) the machine learning method that is chosen. Using biophysical parameters and resampled information over thermal time, our chosen ensemble learning approach explained 42% of the variation in yield, with an associated mean absolute error of 15.2% for samples coming from years not represented in the training data. Comparisons with the literature for field-level corn yield prediction have reported similar accuracies in the United States, but ours requires fewer inputs and less calibration efforts.

CRediT authorship contribution statement

Johann Desloires: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Dino Ienco:** Conceptualization, Writing – review & editing, Supervision, Project administration, Resources, Funding acquisition. **Antoine Botrel:** Investigation, Validation, Supervision, Project administration, Funding acquisition.

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.

Data availability

The data that has been used is confidential

Acknowledgements

The authors would like to thank the European Space Agency (ESA), France for sponsoring the SentinelHub account as part of the Network of Resources (NoR) program, and Syngenta Production&Supply for making its agronomic data available. This research was funded by ANRT, France, grant number CIFRE 2019/1993.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al., 2016. Tensorflow: A system for large-scale machine learning. In: 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16). pp. 265–283.

Angel, J., Widhalm, M., Todey, D., Massey, R., Biehl, L., 2016. The U2U corn growing degree day tool: Tracking corn growth across the US corn belt. Clim. Risk Manag. 15, http://dx.doi.org/10.1016/j.crm.2016.10.002.

Atwell, B., Kriedemann, P., Turnbull, C., 1999. Plants in Action: Adaption in Nature, Performance in Cultivation, first ed. Macmillan Education, Melbourne.

Baret, F., Houlès, V., Guérif, M., 2007. Quantification of plant stress using remote sensing observations and crop models: the case of nitrogen management. J. Exp. Bot. 58 (4), 869–880.

Behmann, J., Mahlein, A.-K., Rumpf, T., Römer, C., Plümer, L., 2015. A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. Precis. Agric. 16 (3), 239–260.

Beres, B.L., Hatfield, J.L., Kirkegaard, J.A., Eigenbrode, S.D., Pan, W.L., Lollato, R.P., Hunt, J.R., Strydhorst, S., Porker, K., Lyon, D., Ransom, J., Wiersma, J., 2020. Toward a better understanding of genotype × environment × management interactions—A global wheat initiative agronomic research strategy. Front. Plant Sci. (ISSN: 1664-462X) 11, http://dx.doi.org/10.3389/fpls.2020.00828, URL https://www.frontiersin.org/articles/10.3389/fpls.2020.00828.

Bertini, F., Brand, O., Carlier, S., Del Bello, U., Drusch, M., Duca, R., Fernandez, V., Ferrario, C., Ferreira, M.H., Isola, C., Kirschner, V., Laberinti, P., Lambert, M., Mandorlo, G., Marcos, P., Martimort, P., Moon, S., Oldeman, P., Palomba, M., Pineiro, J., 2012. Sentinel-2 ESA's optical high-resolution mission for GMES operational services. ESA Bull. Bull. ASE. Eur. Space Agency SP-1322.

Bishop, C.M., Nasrabadi, N.M., 2006. Pattern Recognition and Machine Learning, Vol. 4. Springer.

Boateng, E.Y., Otoo, J., Abaye, D.A., 2020. Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: a review. J. Data Anal. Inf. Process. 8 (4), 341–357.

Bojanowski, J., Musial, J., Sikora, S., Wozniak, E., Dabrowska-Zielińska, K., Slesinski, P., Łaczyński, A., Milewski, T., 2022. Integration of sentinel-3 and MODIS vegetation indices with ERA-5 agro-meteorological indicators for operational crop yield forecasting. Remote Sens. 14, 1238. http://dx.doi.org/10.3390/rs14051238.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Carletto, C., Jolliffe, D., Banerjee, R., 2015. From tragedy to renaissance: improving agricultural data for better policies. J. Dev. Stud. 51 (2), 133–148.

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining. pp. 785–794.

Cisty, M., Soldanova, V., 2018. Flow prediction versus flow simulation using machine learning algorithms. In: Machine Learning and Data Mining in Pattern Recognition: 14th International Conference, MLDM 2018, New York, NY, USA, July 15-19, 2018, Proceedings, Part II 14. Springer, pp. 369–382.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.

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 Sens. Environ. 253, 112174.

Di Salvo, J.I., Lee, C., Salmerón, M., 2021. Regional multi-environment analysis of corn productivity and yield stability as impacted by hybrid maturity. Field Crops Res. 262, 108025. http://dx.doi.org/10.1016/j.fcr.2020.108025.

Duchemin, B.t., Maisongrande, P., Boulet, G., Benhadj, I., 2008. A simple algorithm for yield estimates: Evaluation for semi-arid irrigated winter wheat monitored with green leaf area index. Environ. Model. Softw. 23, 876–892. http://dx.doi.org/10.1016/j.envsoft.2007.10.003.

Durgun, Y.Ö., Gobin, A., Duveiller, G., Tychon, B., 2020. A study on trade-offs between spatial resolution and temporal sampling density for wheat yield estimation using both thermal and calendar time. Int. J. Appl. Earth Obs. Geoinf. 86, 101988.

Duveiller, G., Frederic, B., Defourny, P., 2013. Using thermal time and pixel purity for enhancing biophysical variable time series: An interproduct comparison. IEEE Trans. Geosci. Remote Sens. 51, 2119–2127. http://dx.doi.org/10.1109/TGRS.2012.2226731.

Estes, L., Bradley, B., Beukes, H., Hole, D., Lau, M., Oppenheimer, M., Schulze, R., Tadross, M., Turner, W., 2013. Comparing mechanistic and empirical model projections of crop suitability and productivity: implications for ecological forecasting. Global Ecol. Biogeogr. 22 (8), 1007–1018.

García-Lara, S., Serna-Saldivar, S.O., 2019. Chapter 1 - corn history and culture. In: Serna-Saldivar, S.O. (Ed.), Corn (Third Edition), third ed. AACC International Press, Oxford, ISBN: 978-0-12-811971-6, pp. 1–18. http://dx.doi.org/10.1016/B978-0-12-811971-6.00001-2.

Gbodjo, Y.J.E., Ienco, D., Leroux, L., 2021. Benchmarking statistical modelling approaches with multi-source remote sensing data for millet yield monitoring: a case study of the groundnut basin in central Senegal. Int. J. Remote Sens. 42 (24), 9285–9308.

Gorissen, D., Dhaene, T., De Turck, F., 2009. Evolutionary model type selection for global surrogate modeling. J. Mach. Learn. Res. 10, 2039–2078.

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

Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Vol. 2. Springer.

Hatfield, J., Boote, K., Kimball, B., Ziska, L., Izaurralde, R., Ort, D., Thomson, A., 2011. Climate impacts on agriculture: Implications for crop production. Agron. J. 103, 351–370. http://dx.doi.org/10.2134/agronj2010.0303.

Hauser, L.T., Féret, J.-B., An Binh, N., van der Windt, N., Sil, Â.F., Timmermans, J., Soudzilovskaia, N.A., van Bodegom, P.M., 2021. Towards scalable estimation of plant functional diversity from sentinel-2: In-situ validation in a heterogeneous (semi-)natural landscape. Remote Sens. Environ. (ISSN: 0034-4257) 262, 112505. http://dx.doi.org/10.1016/j.rse.2021.112505.

Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12, 55–67.

Hoffman, L.A., Etienne, X.L., Irwin, S.H., Colino, E.V., Toasa, J.I., 2015. Forecast performance of WASDE price projections for US corn. Agric. Econ. 46 (S1), 157–171.

Holzworth, D.P., Huth, N.I., deVoil, P.G., Zurcher, E.J., Herrmann, N.I., McLean, G., Chenu, K., van Oosterom, E.J., Snow, V., Murphy, C., et al., 2014. APSIM–evolution towards a new generation of agricultural systems simulation. Environ. Model. Softw. 62, 327–350.

Huber, F., Yushchenko, A., Stratmann, B., Steinhage, V., 2022. Extreme gradient boosting for yield estimation compared with deep learning approaches. Comput. Electron. Agric. 202, 107346.

Ines, A.V., Das, N.N., Hansen, J.W., Njoku, E.G., 2013. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. Remote Sens. Environ. (ISSN: 0034-4257) 138, 149–164. http://dx.doi.org/10.1016/j.rse.2013.07.018.

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C., Ustin, S.L., 2009. PROSPECT+SAIL models: A review of use for vegetation characterization. Remote Sens. Environ. (ISSN: 0034-4257) 113, S56–S66. http://dx.doi.org/10.1016/j.rse.2008.01.026, URL https://www.sciencedirect.com/science/article/pii/S0034425709000765. Imaging Spectroscopy Special Issue.

Jonsson, P., Eklundh, L., 2002. Seasonality extraction by function fitting to time-series of satellite sensor data. IEEE Trans. Geosci. Remote Sens. 40 (8), 1824–1832. http://dx.doi.org/10.1109/TGRS.2002.802519.

Kamir, E., Waldner, F., Hochman, Z., 2020. Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. ISPRS J. Photogramm. Remote Sens. (ISSN: 0924-2716) 160, 124–135. http://dx.doi.org/10.1016/j.isprsjprs.2019.11.008.

Kancheva, R., Borisova, D., Georgiev, G., 2007. Spectral predictors of crop development and yield. In: 2007 3rd International Conference on Recent Advances in Space Technologies. pp. 247–251. http://dx.doi.org/10.1109/RAST.2007.4283987.

Kang, Y., Özdoğan, M., 2019. Field-level crop yield mapping with landsat using a hierarchical data assimilation approach. Remote Sens. Environ. 228, 144–163.

Kattenborn, T., Schiefer, F., Frey, J., Feilhauer, H., Mahecha, M., Dormann, C., 2022. Spatially autocorrelated training and validation samples inflate performance assessment of convolutional neural networks. ISPRS Open J. Photogramm. Remote Sens. 5, 100018. http://dx.doi.org/10.1016/j.ophoto.2022.100018.

Kayad, A., Sozzi, M., Gatto, S., Marinello, F., Pirotti, F., 2019. Monitoring within-field variability of corn yield using sentinel-2 and machine learning techniques. Remote Sens. 11 (23), http://dx.doi.org/10.3390/rs11232873.

Khaki, S., Wang, L., Archontoulis, S.V., 2020. A cnn-rnn framework for crop yield prediction. Front. Plant Sci. 10, 1750.

Klompenburg, T., Kassahun, A., Catal, C., 2020. Crop yield prediction using machine learning: A systematic literature review. Comput. Electron. Agric. 177, 105709. http://dx.doi.org/10.1016/j.compag.2020.105709.

Kok, Z.H., Shariff, A.R.M., Alfatni, M.S.M., Khairunniza-Bejo, S., 2021. Support vector machine in precision agriculture: a review. Comput. Electron. Agric. 191, 106546.

Kuwata, K., Shibasaki, R., 2016. Estimating corn yield in the united states with modis evi and machine learning methods. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. 3 (8), 131–136.

Lambert, R., Mansfield, B., Mumm, R., 2014. Effect of leaf area on maize productivity. Maydica 59, 58–63.

Lang, N., Kalischek, N., Armston, J., Dubayah, R., Wegner, J., 2022. Global canopy height regression and uncertainty estimation from GEDI LIDAR waveforms with deep ensembles. Remote Sens. Environ. 268, 112760. http://dx.doi.org/10.1016/j.rse.2021.112760.

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 Sens. 14 (10), 2458.

Lobell, D.B., 2013. The use of satellite data for crop yield gap analysis. Field Crops Res. 143, 56–64.

Lobell, D.B., Thau, D., Seifert, C., Engle, E., Little, B., 2015. A scalable satellite-based crop yield mapper. Remote Sens. Environ. (ISSN: 0034-4257) 164, 324–333. http://dx.doi.org/10.1016/j.rse.2015.04.021.

Lykhovyd, P.V., 2020. Sweet corn yield simulation using normalized difference vegetation index and leaf area index. J. Ecol. Eng. 21, 228–236.

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., Gascon, F., 2017. Sen2Cor for sentinel-2. p. 3. http://dx.doi.org/10.1117/12.2278218.

Marshall, M., Belgiu, M., Boschetti, M., Pepe, M., Stein, A., Nelson, A., 2022. Field-level crop yield estimation with PRISMA and sentinel-2. ISPRS J. Photogramm. Remote Sens. 187, 191–210. http://dx.doi.org/10.1016/j.isprsjprs.2022.03.008.

Marszalek, M., Körner, M., Schmidhalter, U., 2022. Prediction of multi-year winter wheat yields at the field level with satellite and climatological data. Comput. Electron. Agric. 194, 106777.

Mateo-Sanchis, A., Piles, M., Muñoz-Marí, J., Adsuara, J.E., Pérez-Suay, A., Camps-Valls, G., 2019. Synergistic integration of optical and microwave satellite data for crop yield estimation. Remote Sens. Environ. (ISSN: 0034-4257) 234, 111460. http://dx.doi.org/10.1016/j.rse.2019.111460.

Mishra, S., Mishra, D., Santra, G.H., et al., 2016. Applications of machine learning techniques in agricultural crop production: a review paper. Indian J. Sci. Technol. 9 (38), 1–14.

Morain, S., Williams, D., 1975. Wheat production estimates using satellite images 1. Agron. J. 67 (3), 361–364.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. ISPRS J. Photogramm. Remote Sens. 66, 247–259. http://dx.doi.org/10.1016/j.isprsjprs.2010.11.001.

Ngoune Tandzi, L., Mutengwa, C.S., 2020. Estimation of Maize (Zea mays L.) yield Per Harvest Area: Appropriate methods. Agronomy (ISSN: 2073-4395) 10 (1), http://dx.doi.org/10.3390/agronomy10010029.

Nleya, T., Chungu, C., Kleinjan, J., 2019. Corn growth and development. pp. 5–8.

Novelli, F., Spiegel, H., Sandén, T., Vuolo, F., 2019. Assimilation of sentinel-2 leaf area index data into a physically-based crop growth model for yield estimation. Agronomy (ISSN: 2073-4395) 9 (5), http://dx.doi.org/10.3390/agronomy9050255.

Palagyi, A., Nemeth, J., 1996. Maize hybrid seed production by the mutual random mating of the parental components. Cereal Res. Commun. 307–316.

Palosuo, T., Kersebaum, K.C., Angulo, C., Hlavinka, P., Moriondo, M., Olesen, J.r.E., Patil, R.H., Ruget, F., Rumbaur, C., Takáč, J., et al., 2011. Simulation of winter wheat yield and its variability in different climates of europe: A comparison of eight crop growth models. Eur. J. Agron. 35 (3), 103–114.

Pasqualotto, N., D'Urso, G., Bolognesi, S.F., Belfiore, O.R., Van Wittenberghe, S., Delegido, J., Pezzola, A., Winschel, C., Moreno, J., 2019. Retrieval of evapotranspiration from sentinel-2: Comparison of vegetation indices, semi-empirical models and SNAP biophysical processor approach. Agronomy 9 (10), http://dx.doi.org/10.3390/agronomy9100663.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine learning in python. J. Mach. Learn. Res. 12 (Oct), 2825–2830.

Rembold, F., Atzberger, C., Savin, I., Rojas, O., 2013. Using low resolution satellite imagery for yield prediction and yield anomaly detection. Remote Sens. (ISSN: 2072-4292) 5 (4), 1704–1733. http://dx.doi.org/10.3390/rs5041704.

Rudorff, B., Batista, G., 1990. Spectral response of wheat and its relationship to agronomic variables in the tropical region. Remote Sens. Environ. 31 (1), 53–63.

Runge, E., 1968. Effects of rainfall and temperature interactions during the growing season on corn yield 1. Agron. J. 60 (5), 503–507.

Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36, 1627–1639.

Schauberger, B., Archontoulis, S., Arneth, A., Balkovič, J., Ciais, P., Deryng, D., Elliott, J., Folberth, C., Khabarov, N., Müller, C., Pugh, T., Rolinski, S., Schaphoff, S., Schmid, E., Wang, X., Schlenker, W., Frieler, K., 2017. Consistent negative response of US crops to high temperatures in observations and crop models. Nature Commun. 8, 13931. http://dx.doi.org/10.1038/ncomms13931.

Schwalbert, R.A., Amado, T.J., Nieto, L., Varela, S., Corassa, G.M., Horbe, T.A., Rice, C.W., Peralta, N.R., Ciampitti, I.A., 2018. Forecasting maize yield at field scale based on high-resolution satellite imagery. Biosyst. Eng. (ISSN: 1537-5110) 171, 179–192. http://dx.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. Int. J. Appl. Earth Obs. Geoinf. (ISSN: 1569-8432) 107, 102697.

Shao, Y., Lunetta, R., Wheeler, B., Iiames, J., Campbell, J., 2016. An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data. Remote Sens. Environ. 174, 258–265. http://dx.doi.org/10.1016/j.rse.2015.12.023.

Sinergise, 2019. Eo-learn. https://github.com/sentinel-hub/eo-learn.

Skakun, S., Vermote, E., Franch, B., Roger, J.-C., Kussul, N., Ju, J., Masek, J., 2019. Winter wheat yield assessment from landsat 8 and sentinel-2 data: Incorporating surface reflectance, through phenological fitting, into regression yield models. Remote Sens. 11 (15), 1768.

Stone, P., Wilson, D., Jamieson, P., Gillespie, R., 2000. Water deficit effects on sweet corn. II. Canopy development. Crop Pasture Sci. 52, 115–126. http://dx.doi.org/10.1071/AR99145.

Tucker, C.J., Holben, B., Elgin, Jr., J., Mcmurtrey III, J., 1980. Relationship of spectral data to grain yield variation. Photogramm. Eng. Remote Sens. 46.

Van Klompenburg, T., Kassahun, A., Catal, C., 2020. Crop yield prediction using machine learning: A systematic literature review. Comput. Electron. Agric. 177, 105709.

Verger, A., Filella, I., Baret, F., Peñuelas, J., 2016. Vegetation baseline phenology from kilometric global LAI satellite products. Remote Sens. Environ. (ISSN: 0034-4257) 178, 1–14. http://dx.doi.org/10.1016/j.rse.2016.02.057.

Wang, M., lu Tao, F., Shi, W.-J., 2014. Corn yield forecasting in northeast China using remotely sensed spectral indices and crop phenology metrics. J. Integr. Agric. (ISSN: 2095-3119) 13 (7), 1538–1545. http://dx.doi.org/10.1016/S2095-3119(14)60817-0.

Wang, W., Yao, X., Yao, X., Tian, Y., Liu, X., Ni, J., Cao, W., Zhu, Y., 2012. Estimating leaf nitrogen concentration with three-band vegetation indices in rice and wheat. Field Crops Res. (ISSN: 0378-4290) 129, 90–98. http://dx.doi.org/10.1016/j.fcr.2012.01.014, URL https://www.sciencedirect.com/science/article/pii/S0378429012000299.

Weiss, M., Baret, F., 2016. S2ToolBox Level 2 Products: LAI, FAPAR. FCOVER, Institut National de la Recherche Agronomique (INRA), Avignon.

Westcott, P., Jewison, M., 2014. Weather effects on expected corn and soybean yields. pp. 73–105.

Whitcraft, A.K., Becker-Reshef, I., Justice, C.O., 2015. A framework for defining spatially explicit earth observation requirements for a global agricultural monitoring initiative (GEOGLAM). Remote Sens. 7 (2), 1461–1481.

Xu, X., Gao, P., Zhu, X., Guo, W., Ding, J., Li, C., Zhu, M., Wu, X., 2019. Design of an integrated climatic assessment indicator (ICAI) for wheat production: a case study in Jiangsu province, China. Ecol. Indic. 101, 943–953.

You, J., Li, X., Low, M., Lobell, D., Ermon, S., 2017. Deep gaussian process for crop yield prediction based on remote sensing data. In: Thirty-First AAAI Conference on Artificial Intelligence.