

Article

Assessing Durum Wheat Yield through Sentinel-2 Imagery: A Machine Learning Approach

Maria Bebie, Chris Cavalaris and Aris Kyparissis *

Department of Agriculture Crop Production and Rural Environment, University of Thessaly, Fytokou Str., 38446 Volos, Greece

* Correspondence: akypar@uth.gr

Abstract: Two modeling approaches for the estimation of durum wheat yield based on Sentinel-2 data are presented for 66 fields across three growing periods. In the first approach, a previously developed multiple linear regression model (VI-MLR) based on vegetation indices (EVI, NMDI) was used. In the second approach, the reflectance data of all Sentinel-2 bands for several dates during the growth periods were used as input parameters in three machine learning model algorithms, i.e., random forest (RF), k-nearest neighbors (KNN), and boosting regressions (BR). Modeling results were examined against yield data collected by a combine harvester equipped with a yield mapping system. VI-MLR showed a moderate performance with $R^2 = 0.532$ and $RMSE = 847 \text{ kg ha}^{-1}$. All machine learning approaches enhanced model accuracy when all images during the growing periods were used, especially RF and KNN ($R^2 > 0.91$, $RMSE < 360 \text{ kg ha}^{-1}$). Additionally, RF and KNN accuracy remained high ($R^2 > 0.87$, $RMSE < 455 \text{ kg ha}^{-1}$) when images from the start of the growing period until March, i.e., three months before harvest, were used, indicating the high suitability of machine learning on Sentinel-2 data for early yield prediction of durum wheat, information considered essential for precision agriculture applications.

Keywords: durum wheat; yield modeling; Sentinel-2; machine learning; vegetation indices

Citation: Bebie, M.; Cavalaris, C.; Kyparissis, A. Assessing Durum Wheat Yield through Sentinel-2 Imagery: A Machine Learning Approach. *Remote Sens.* **2022**, *14*, 3880. <https://doi.org/10.3390/rs14163880>

Academic Editor: Stefano Dietrich

Received: 1 July 2022

Accepted: 9 August 2022

Published: 10 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The early, accurate and broadscale prediction of durum wheat yield has always been a great challenge for the scientific community, who aim to provide vital information for farmers, the food industry, policymakers, and other stakeholders [1]. Each one anticipates considerable assets from such a forecast. Farmers, for instance, may empower their decisions by aligning the inputs according to the expected production and schedule precision farming applications. Food industry and commodity traders are interested on early estimates of regional productivity to arrange imports or exports, manage logistics, and establish price policies. National and international bodies are concerned about country level, down to field-scale productivity, to gather crucial information that may assist to direct agriculture subsidies, identify yield-limiting and yield-enhancing areas and promote the sustainable use of natural resources by establishing land management plans. National and private crop insurance entities may also utilize such early information to provide effective crop insurance plans.

The first yield prediction models, e.g., CERES [2], CROPSYST [3], and SAFY [4], were empirical statistical and process-oriented approaches that aimed to simulate crop growth based on numerous biophysical, environmental, and managerial parameters that are often hardly accessible or require great effort, especially for broadscale or high resolution predictions [1,5].

With the rapid development of remote sensing (RS) technology, broadscale data with high spatial and temporal resolution have been made widely available. Remote sensing

data obtained from the National Oceanic and Atmospheric Administration's (NOAA) Advanced Very High Resolution Radiometer (AVHRR) instrument have been used to monitor large-scale cropping systems and to forecast yield from the 1980s [1]. Since then, multispectral instruments onboard various satellites, such as LANDSAT [6], MODIS [7], SPOT [8], and lately the Sentinel satellites launched by the Copernicus program of the European Space Agency (ESA) have provided free, independent, and continued global remote sensing observations for modeling crop growth [9–11]. The new RS-based models are either empirical linear regressions between biophysical parameters and crop yield or biomass production models. The first may incorporate above-ground crop parameters (biomass, LAI, chlorophyll etc.) accessed through vegetation indices (VIs), such as the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Soil-Adjusted Vegetation Index (SAVI) etc. [12–15]. The latter may also incorporate environmental data like topography, weather information, fraction of absorbed photosynthetically active radiation (fAPAR), and other parameters from common simulation models to estimate crop productivity and the resulting final yield [11,16,17]. The first method does not provide an explanation of physiological processes, dismissing important components of crop yield, while the latter method is complicated and may require some hardly accessible inputs [18]. The compromises in both approaches result in moderate prediction power, with a coefficient of determination (R^2) ranging generally from 0.50 to 0.75 and models that are usually applicable at a regional scale [15,19,20].

An important technological breakthrough to overcome the above limitations has been introduced lately, with the wide implementation of machine learning (ML) techniques over broadscale RS data [21]. These techniques can handle multi-dimensional datasets [22] and thus are capable to deal with complexities in yield prediction by exploiting the underlying physiological processes and interactions hidden by the simple linear regression models [21,23]. With the ability to manage both linear and non-linear relationships [24] and potential to exploit the most important relations, ML can utilize the entire range of the primary spectral information instead of a set of few spectral bands used in VIs [21]. Hence, ML is capable to extract the most from the RS information, while correctness, modularity, and reusability are feasible over pre-established baseline workflows, allowing transfer to other crops and locations [25].

There are numerous ML techniques utilized to represent various biophysical processes like artificial neural networks (ANNs) [21], support vector machine (SVM) [26], Gaussian process regression (GPR) [5], partial least squares regression (PLSR) [26], multi-layer perceptron (MLP) [26], random forest (RF) [27], k-nearest neighbors (KNN) [26], boosting regressions (BR) [27], and others [28]. Each technique has its strengths and drawbacks, making it suitable for different purposes. ANNs, for instance, present high accuracy and stability and offer a higher computational speed, while support vector regression (SVR), an extension of SVM, presents good intrinsic generalization ability and has a robustness to noise in the case of limited availability of the reference samples [28]. RF has been broadly used as a classification tool and may handle efficiently large datasets, while it is less prone to outliers and reduces the risk of overfitting [5,29]. KNN is noise-tolerant but performs poorly when applied to small training datasets [30]. BR is capable of squeezing additional predictive accuracy out of other ML algorithms but increases the model's complexity [27].

ML algorithms have been used recently to exploit complex relationships between RS information and various crop traits. Their ability to handle both linear and non-linear relations has made them suitable for the investigation of the non-linear traits of RS and crop yield [31]. Moreover, they are flexible to decide which bands are most valuable for predicting crop yield [11]. Since ML algorithms require large training and validation datasets to build accurate models, an important precondition is the availability of an adequate number of yield data at a resolution similar to the RS information. This has also been made possible recently by fixing yield mapping devices to the harvesting equipment. Recent studies have attempted to match these yield data with RS spectral information to feed ML

algorithms and develop yield estimation models. In [11], for instance, RF models were trained and validated using over 8000 data points from yield monitors onboard combine harvesters from 39 wheat fields in the UK. In [32], the authors matched several VIs retrieved from the Sentinel-2 satellites with wheat grain data retrieved from a combine harvester. Subsequently, they implemented different ML techniques to model the within-field grain variability. The authors tested also different time intervals and single dates during the growing period to identify the most suitable model feeding the dataset. Their findings suggest that the RF algorithm was the most suitable for estimating wheat yield. The coefficient of determination reached an $R^2 = 0.89$ value when all available images were used, but since the study was conducted in northern Spain, only two cloud-free images were available per year. Cloud cover and scarcity of available satellite images is a serious drawback for RS modeling in high latitudes [11]. Nevertheless, in [32], RF with only one image at stem elongation posed a relatively high R^2 , offering potential for precision farming management applications. In [22], three separate empirical RF models were created based on pre-sowing, mid-season and late-season conditions to explore the time changes in the predictive ability of the model. The models concerned wheat, barley and canola and the data were aggregated at a field-size spatial resolution. As more within-season information became available, the models performed better. Han et al. [5] also agree with this suggestion. It appears that RF models are more suitable for the whole growing period, while other ML methods, like ANN and KNN, are more appropriate for specific growth periods [26]. Han et al. [5] separated the growth period of wheat into four time windows and assessed their corresponding predictive ability by testing eight ML algorithms. They found that a time window including the whole growing period (starting from sowing at end of September and reaching the harvest period in June) provided the best prediction accuracy. SVM, GPR, and RF represent the three best methods for predicting wheat yields, with an R^2 of about 0.8. SVM performed slightly better for the whole growing period.

Using explainable predictors in relation to crop growth and development, combining agronomic features with RS information is a valuable step in building ML baselines for large-scale crop yield forecasting [25,33]. Leaf area index (LAI), for instance, improved significantly an RF modeling approach for predicting within field wheat yield [32]. Leaf angle distribution (LAD) and leaf mean tilt angle (MTA) are some other important canopy structural traits. They are used to quantify the direction of the leaf surface, and combined with LAI, they help determine the light interception [26]. In [26], the authors found that the red edge band of Sentinel-2 had a strong correlation with MTA. Consequently, they applied three different ML techniques for estimating the three canopy structural parameters and found that RF had the best association with findings from the PROSAIL model simulations and MLP with actual field data. Filippi et al. [22] suggest a workflow collating in a data cube high-density data that varies only in space (e.g., gamma radiometrics), high-density data that varies in space and time (e.g., in-season imagery), and lower-resolution data that varies in space (e.g., soil physical test results) or space and time (rainfall). Within-season variables proved to be vital in the models, with received rainfall, forecasted rainfall, and within-season EVI images amongst the most important covariates.

In a previous study in Greece [34], we obtained EVI and NDVI from Sentinel-2 and proved that the two VIs presented a high correlation with in situ measurements of wheat canopy reflectance and were good predictors of LAI. It was also shown that the Normalized Multiband Drought Index (NMDI) corresponded very well to in situ measurements of leaf water potential. The same index presented a very good correlation with soil apparent electric conductivity (ECa) when it was retrieved from an image of bare soil (right before sowing). The above findings were used to build linear regression models that relied strictly on RS data to address explainable wheat yield predictors. The models utilized EVI and NDVI as plant vegetation signals, NMDI as plant water signal, and bare soil NMDI as a soil signal. Accordingly, the models were evaluated on 31 durum wheat fields for two growing periods and for different time intervals as well as single dates, finding that the best period for predicting wheat yield was the time window from 20 April to 31 May (R^2

= 0.629, RMSE = 528 kg ha⁻¹), which is rather late because it is very close to harvest. Nevertheless, the last 10 days of April also provided high prediction accuracy ($R^2 = 0.587$, RMSE = 568 kg ha⁻¹) more than 1 month earlier. The proposed model relies on easily accessible Sentinel-2 data of high spatial resolution that are extremely valuable for the small farms of Southern Europe, but the optimum period is still rather late for precision farming applications. On the other hand, ML approaches have the potential to explore hidden information on the datasets, providing the opportunity for early season estimations. However, the relevant research presented above is still rather limited and has some serious gaps, mainly on the number of available cloud-free images.

In the present work, we introduce some ML techniques to a full-season Sentinel-2 dataset and compare the performance with our previous linear regression models. The aim is to increase the accuracy and improve the timeliness of wheat yield prediction by retaining a simple database restricted to the high spatial resolution of the Sentinel-2 images. In this approach, we avoid estimating VIs and use all the available Sentinel-2 bands, allowing the ML algorithms to decide what is the more valuable information. We also introduce extra yield data from 66 fields across three growing periods and check again for different time intervals to find the optimum predicting period.

2. Materials and Methods

2.1. Study Sites

All study sites are located in Thessaly plain, central Greece, and the monitoring involved the 2017–2018, 2018–2019, and 2019–2020 growing periods (Figure 1 and Table 1). In sum, 66 fields—21 from the 2017–2018 growing period, 21 from 2018–2019, and 24 from 2019–2020—were used. The criteria for selecting the fields were i) the availability of yield map data and ii) an adequate size, scheme, and orientation capable of providing a sufficient number of Sentinel-2, 10 × 10 m resolution pixels. Fields were cultivated with durum wheat of different varieties (Iridae, Meridiano, Normano, Simeto, Svevo) and belonged to different farmers who followed their own cultivation practices. That way, the results of the study could be evaluated at different conditions, ensuring that they are applicable regardless of the cropping practices. For all fields, sowing took place during November and harvesting during June.

Table 1. Details of the studied fields.

Growing Period	No of Fields	Area, ha	No of Pixels	No of Images
2017–2018	21	53.04	5304	35
2018–2019	21	50.31	5031	26
2019–2020	24	85.91	8591	32
Total	66	189.26	18,926	93

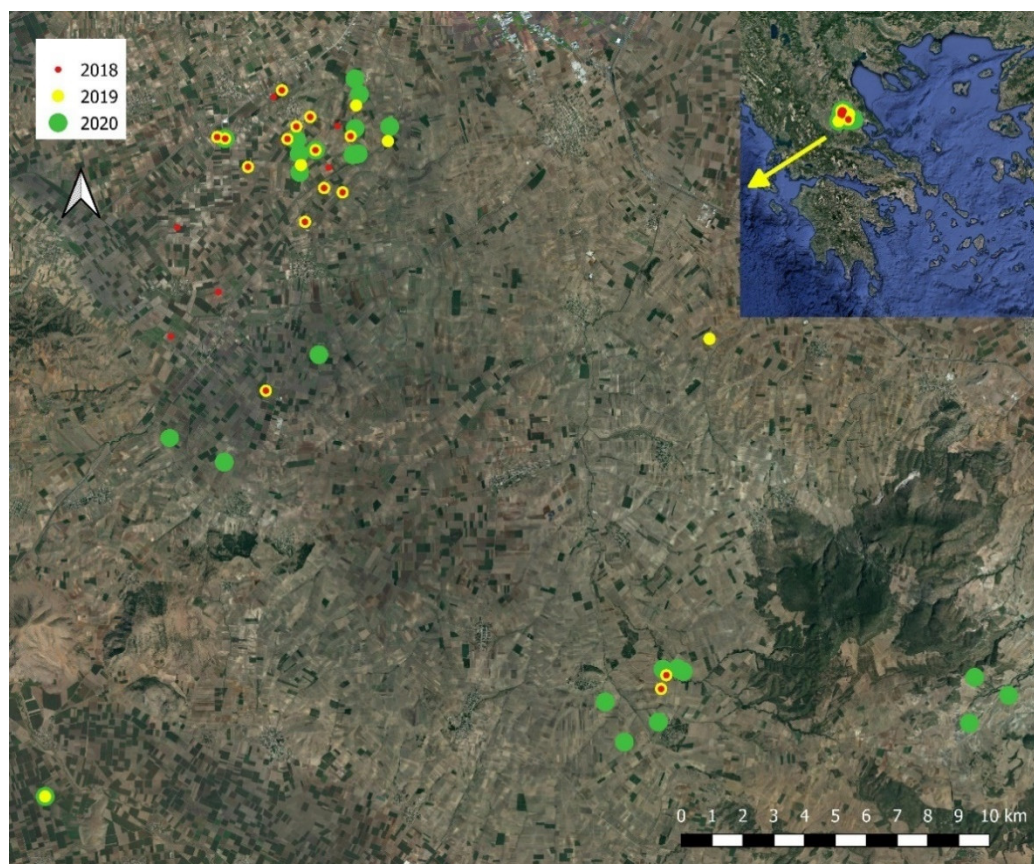


Figure 1. Satellite image of the Thessaly plain, Greece (inset), with the studied fields indicated with different colors for the three growing periods.

2.2. Yield Measurement

At the end of the growing periods (June) the fields were harvested with a John Deere S660i combine harvester equipped with a yield mapping system providing through the associated MyJD software [35] (<https://myjohndeere.deere.com>, accessed on 24 October 2021), yield maps in a point vector format at a spatial resolution of 1.75 by approximately 2.5 m (depending on the traveling speed). The initial maps were further processed manually in QGIS [36] for the removal of outliers (due to start and end point grain flow delays). Accordingly, the yield maps were interpolated to rasters by the inverse distance weighting (IDW) process of QGIS and finally, resampled at 10×10 m pixel size corresponding to the Sentinel-2 image pixels (see also [34]).

2.3. Satellite Data

A total of 93 cloud-free Sentinel-2 (A and B) images from October to June across the study period (2017–2020, Table 1) were downloaded from ESA's Copernicus Open Access Hub [37] (<https://scihub.copernicus.eu/>, accessed on 26 October 2021). During the 3-year study, there was no period longer than 20 days with no images available due to cloud cover. The MultiSpectral Instruments (MSI), onboard the Sentinel-2 satellites, provide information at 13 spectral bands (443–2190 nm), at a variable spatial resolution of 10, 20 or 60 m pixel size and with 5-day revisit time. In the present study, Level 2A (radiometrically and atmospherically corrected) bottom of atmosphere (BOA) reflectance products provided by ESA, were used. For all images, all bands were resampled at 10 m pixel size using the SNAP—ESA Sentinels Application Platform version 7.0 [38] (<http://step.esa.int>, accessed on 26 October 2021) free open-source software. The images included a total of 18,926 pixels from 66 fields during a 3-year period, and spectral information comprising 13 spectral bands per pixel was extracted on a tabular database.

2.4. Modeling

Two modeling approaches were used for the estimation of wheat yield, a previously developed multiple linear regression approach using vegetation indices as yield predictor (VI-MLR) and a machine learning (ML) approach exploring the performance of three ML algorithms on the Sentinel-2 MSI multiband dataset as yield predictors.

2.4.1. VI-MLR-Based Model

In the first approach, we adopt the best performing multiple linear regression model of our previous study [35]. This model involves the EVI integral from April 20 to May 31 as plant signal, NMDI before sowing as soil signal and NMDI at the end of April as water signal (independent variables). For the scope, the B2 (R_{490}), B4 (R_{665}), B8 (R_{842}), B11 (R_{1610}), B12 (R_{2190}) Sentinel-2 spectral bands of the 18,926 pixels were used to estimate corresponding time series of EVI and NMDI as follows:

$$\text{Enhanced Vegetation Index, EVI} = 2.5 \frac{R_{842} - R_{665}}{R_{842} + 6R_{665} - 7.5R_{490} + 1} \quad (1)$$

$$\text{Normalized Multiband Drought Index, NMDI} = \frac{R_{842} - (R_{1610} - R_{2190})}{R_{842} + (R_{1610} - R_{2190})} \quad (2)$$

where R_x , reflectance at wavelength x , with x denoting the center wavelength of the corresponding Sentinel-2 band. Accordingly, the time series were linearly interpolated without any smoothing process producing complete daily datasets for the interested period (April 20 to May 31 plus one image from October, prior sowing).

2.4.2. ML-Based Models

In the second approach, the reflectance data for all Sentinel-2 spectral bands were used as independent variables in three commonly used ML algorithms, i.e., random forest (RF), k-nearest neighbors (KNN) and boosting regression (BR).

Random forest was chosen for its suitability in yield prediction by exploring important underlying information from the whole growing period while reducing the risk of overfitting [5,26,29]. RF is a supervised ML technique that establishes decision trees on different subsets of a dataset. Each tree is a predictor built by selecting a random sample of the original dataset, but all the trees in the forest have the same distribution characteristics. After generating a large number of individual trees, the algorithm will choose the most popular classes based on the majority votes of the predictors [5].

K-nearest neighbors is also a supervised ML technique that is used to solve both classification and regression problems. It is a non-parametric algorithm not making any assumption on underlying information and it is considered noise-tolerant and suitable to focus on more specific periods [26,30]. It relies on an instance-based learning concept by assuming the similarity between a new predictor variable and a training group. Then it classifies the new predictor into the most similar category. The classification depends greatly on the distance of the predictor variables to the nearest training group [5,30].

Boosting regression is an ML generic algorithm with an enhanced prediction accuracy [28,39] that relies on a family of single ML techniques. BR assumes that all the single predictions of the ML algorithms are weak and performs parallel computations following an iterative process using average and weighted average estimations to vote for the most dominant prediction. After many iterations, the boosting algorithm combines these weak predictions into a single strong prediction rule.

From the 93 images spanning across the three growing periods, 33 images between the end of October and the end of May were selected. This image selection resulted in a dataset with 11 images per growing period (one to three images per month), with dates no more than ± 5 days between growing periods. For each date, the 13 Sentinel-2 bands were used as independent variables in the ML algorithms, resulting in 143 variables in the full dataset (13 bands \times 11 dates). Initially, all images (dates) were used and gradually—

one date at a time—starting from the latest one, was excluded, to find the earliest time interval with sufficient prediction accuracy.

In all modeling approaches, data were randomly split into a training set (50% of data) and a validation set (50% of data) for assessment of performance efficiency. The same data splitting was used for all models.

2.5. Statistics

The multiple linear regressions between the independent variables and final yield for the first modeling approach and the machine learning for the second, were performed with JASP software version 0.16 [40]. Models' performance was evaluated by the coefficient of determination (R^2), root mean square error (RMSE) and slope of the best-fit line between measured and modeled yield values.

3. Results

The performance of the first VI-MLR modeling approach was medium with $R^2 = 0.532$ and $RMSE = 847 \text{ kg ha}^{-1}$ (Figure 2a). As shown in Figure 2a, yield for several fields especially during the 2018–2019 growing period (red points on top left and right) are not well predicted by the model, deteriorating its performance. As a result, the slope of the best-fit line is 0.536, far from 1 which would indicate a perfect fit.

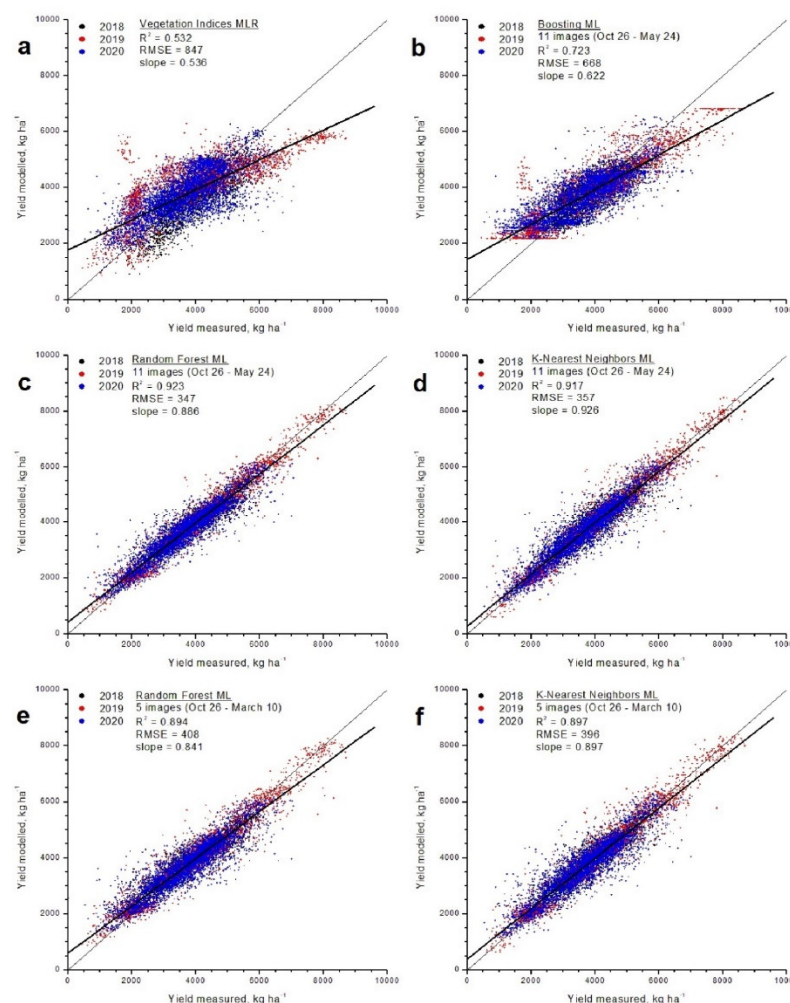


Figure 2. Relationships between measured and modeled yield based on the vegetation indices approach through multiple linear regression (VI-MLR, a) and machine learning approaches (ML, b–f). The type of regression, the number of images per growing period used in ML regressions with

the corresponding date span, the coefficient of determination (R^2), the root mean square error (RMSE) and the slope of the best-fit line are shown in the inserts. Data concern 66 fields during three growing periods (indicated by different colors), corresponding to 9463 pixels. The thin black line corresponds to the 1:1 line and the thick black line to the best-fit line.

In the second modeling approach, all three ML algorithms performed better than the VI-MLR model, when images from the start of the growing period until the end of April or later are used (33 images in the full dataset, 11 images per growing period). However, even though BR performs better than VI-MLR ($R^2 = 0.723$) it retains rather high RMSE (668 kg ha⁻¹) and low slope (0.622) (Table 2 and Figure 2b). Interestingly, the same fields that are not well predicted by the VI-MLR approach are also not well predicted by the BR model. RF and KNN, however, show very good and similar performances, with $R^2 > 0.91$, RMSE < 360 and slopes close to 1 (Figure 2c,d). In both cases, all fields appear very close to the 1:1 line without any outliers. Additionally, even though BR loses its superiority to VI-MLR when images before the end of April are used, both RF and KNN retain their high performance, even when less than five or only three images per growing period, corresponding to dates until mid-March or mid-January respectively, are used (Figure 2e,f and Table 2).

Table 2. Performance comparison of the three machine learning approaches for yield estimation, with the use of different number of images per growing period. R^2 , coefficient of determination and RMSE, root mean square error (kg ha⁻¹) between measured and modeled yield. Data concern 66 fields during three growing periods, corresponding to 9463 pixels.

Dates	Random Forest (RF)		K-Nearest Neighbors (KNN)		Boosting Regression (BR)	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
26 October–24 May 11 images	0.923	347	0.917	357	0.723	668
26 October–29 April 8 images	0.915	366	0.908	375	0.684	709
26 October–10 March 5 images	0.894	408	0.897	396	0.460	938
26 October–28 February 4 images	0.890	419	0.897	397	0.410	980
26 October–19 January 3 images	0.871	455	0.883	425	0.357	1009

4. Discussion

Utilizing satellite RS data to build yield prediction models has become a common approach since such data have been made publicly available. The high spatial resolution and short revisit time of the ESA Sentinel-2 satellites has provided extra interest in the above approach and several attempts have been made recently to estimate or predict wheat yield in small fields through RS information. However, relevant literature is still rather scarce, mainly due to low availability of spatial yield data produced by combine harvesters with yield mapping systems [11,18,32,34]. Large amounts of such data are essential for building and validating the models.

In this paper, we evaluated two different modeling approaches for the estimation of durum wheat yield from Sentinel-2 satellite data. In the first VI-MLR approach, vegetation indices are used in multiple linear regressions with yield, while the second approach concerns machine learning with three different algorithms, random forest (RF), k-nearest neighbors (KNN) and boosting regressions (BR).

Clearly, machine learning with RF and KNN showed far superior performance compared to VI-MLR or BR. Our results agree with the findings in [11,29,32], where RF regression models for predicting wheat yield outperformed multiple linear regression models. RF

models have key advantages over traditional regression models for yield estimation, because they explore relationships between explanatory variables to control for confounding factors [11]. They separate a random subset from calibration for performance testing and use only the remaining set of data for model training, reserving thus information for assessing model accuracy [29]. In [5] the authors compared the performance of eight ML algorithms in predicting wheat yield from RS data and demonstrated also that RF was the best method. The worst performance, however, was found for KNN. KNN presented an R^2 of about 0.65 between the predicted and the observed wheat yields, and an RMSE over 1000 kg ha^{-1} , that the authors consider insufficient for wheat yield prediction. These findings are in contrast with our findings where KNN performed comparable to RF with an R^2 of 0.917 for the whole growing period and an RMSE of 357 kg ha^{-1} . The authors in [5], however, derived RS data from the MODIS satellite instrument, which has a low spatial resolution of 250 m, and KNN algorithm relies on the distance of the predictor variables to the nearest training group known to the model [41], so the low spatial resolution may have provided some irrelevant or biased neighbor values. According to [31], the KNN model is very sensitive to the selection of the k value: increasing the k reduces the variance, but may increase the bias. In [31], KNN performed very well for predicting wheat yield, but the authors used measurements of whole field yields as training and validation datasets. It is also worth mentioning that BR presented the worst performance, even though the algorithm is considered to derive extra power compared to other ML techniques [39,42].

Our modeling approaches suggest a relatively simple workflow, since they are based on raw Level 2A reflectance data, i.e., without secondary level estimation of vegetation indices and neither use any additional biophysical or meteorological parameters. Adding secondary level Vis to the basic Sentinel-2 dataset has not always shown an improvement in wheat yield estimation [11]. According to the authors in [11], it is possible to produce accurate maps of within-field yield variation at 10 m resolution (RMSE 660 kg ha^{-1}) using only Sentinel-2 raw data. In [32] the authors assessed the biophysical parameter of LAI with the PROSAIL radiative transfer model, but the improvement in wheat prediction was rather small in contrast with Sentinel-2 basic bands and VIs.

In order to identify the best period for yield prediction and explore opportunities for early yield estimations, we split our 11 images per growing period dataset into smaller periods down to single dates. The availability of a high number of cloud-free Sentinel-2 images in Southern Europe allowed us to fine tune the models by investigating very small timeframes even down to 5 days which is the Sentinel-2 revisiting period. Our results reveal that even when only a few images from the start to the middle of the growing period are used, the accuracy of prediction remains very high. The lowest RMSE for the RF method was 347 kg ha^{-1} when the whole growing period was accounted and increased to 419 kg ha^{-1} for the period of sowing until end of February, an error that is still acceptable. Other studies have also tried to identify optimum period for wheat yield prediction by RS data. In [11] for instance, the authors found that the accuracy of the estimation increases considerably when additional RS information from December to June is provided, but the findings rely only on three cloud-free available images since the study was conducted in the UK. In [5] it is shown that ML models based on MODIS RS data can accurately predict yield 1~2 months before the harvesting dates at a county level in China. Sentinel data in [32] posed that single date images at stem elongation can provide good estimations of wheat yield by using an RF model. Nevertheless, this study relied also in a couple of satellite images per year. Our results explore the whole growing period in detail and demonstrate clearly that it is feasible to predict durum wheat yield in Southern Europe with a high accuracy as early as January. Such information would be extremely important for making management decisions for the whole field, as also for precision agriculture applications. Variable rate application of nitrogen in wheat during the springtime dressings is an essential practice for improving fertilizer efficiency, optimizing inputs, reducing risks of N leaching, and ensuring N applications according to defined legislation limits [43,44]. The concept for deciding the amount of fertilizer to be applied at different management

zones is greatly based on spatial predictions of expected yields [44] and satellite available data at a high resolution are extremely valuable for that purpose [45]. According to [46] Sentinel-2 imagery was successful for delineation of management zones after Zadocks growth stage 30, and thus is useful for producing fertilization maps for the upcoming season. In [45] it is denoted that satellite data best represented nitrogen uptake in BBCH 39 and 55 growth stages (BBCH scale is the same as the Zadocks scale). The stages from 30 to 39 describe the period from plant pseudo stem erection to just visible flag leaf. These are the stages when springtime dressings of nitrogen are applied and in Southern Europe occur from end of February to the beginning of April. Providing a yield estimation as early as February, as depicted in our study, by utilizing primary Level2A data from Sentinel-2 images may give a new dimension to such precision agriculture applications.

Even though the results of this study demonstrated that machine learning techniques are promising for yield estimation/prediction, they have to be extended in space and time (more growing periods) and for different crops to be generally applicable. All fields used in the study are located in the same plain in close vicinity, with maximum between-field distance approximately 30 km, i.e., meteorological conditions are similar between fields. Consequently, an obvious next step would be to examine the performance of our modeling approach for fields of different areas and during more growing periods (i.e., under different climatic conditions), incorporating meteorological parameters as independent variables in the process, as proposed also by [11]. However, as the number of data may enormously increase, classical ML techniques like RF and KNN may reach their limits. In that case, deep learning (DL) architectures that are capable to process unstructured data at maximum capacity and explore more subtle dependencies may be the solution [47,48]. DL is a branch of ML that has come to the fore in natural language processing and image classification and have taken the lead when it comes to image-based analysis [48].

5. Conclusions

In this study, three machine learning algorithms were used for the estimation of durum wheat yield based on Sentinel-2 satellite data and compared to a previously developed multiple linear regression model based on vegetation indices (VI-MLR). Modeling results were examined against yield data collected by a combine harvester equipped with a yield mapping system. All machine learning approaches showed enhanced estimation accuracy compared to VI-MLR, when all images during the growing periods were used, especially random forest and k-nearest neighbors. Additionally, RF and KNN accuracy remained high when images from the start of the growing period until March, i.e., 3 months before harvest, were used, indicating the high suitability of machine learning on Sentinel-2 data for early yield prediction of durum wheat, essential information for precision agriculture applications.

Author Contributions: Conceptualization, C.C. and A.K.; investigation, M.B., C.C. and A.K.; writing—original draft preparation, C.C. and A.K.; writing—review and editing, C.C. and A.K.; visualization, M.B. and A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data are available from the authors upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Basso, B.; Cammarano, D.; Carfagna, E. Review of crop yield forecasting methods and early warning systems. In *Proceedings of the Proceedings of the First Meeting of the Scientific Advisory Committee of the Global Strategy to Improve Agricultural and Rural Statistics*; FAO Headquarters, Rome, Italy, 2013; Volume 18, p. 19.
2. Ritchie, J.T.; Otter, S. Description and performance of CERES-Wheat: A user-oriented wheat yield model. *USDA-ARS* **1985**, *38*, 159–175.

3. Van Evert, F.K.; Campbell, G.S. CropSyst: A collection of object-oriented simulation models of agricultural systems. *Agron. J.* **1994**, *86*, 325–331.
4. Duchemin, B.; Maisongrande, P.; Boulet, G.; Benhadj, I. A simple algorithm for yield estimates: Evaluation for semi-arid irrigated winter wheat monitored with green leaf area index. *Environ. Model. Softw.* **2008**, *23*, 876–892. <https://doi.org/10.1016/j.envsoft.2007.10.003>.
5. Han, J.; Zhang, Z.; Cao, J.; Luo, Y.; Zhang, L.; Li, Z.; Zhang, J. Prediction of winter wheat yield based on multi-source data and machine learning in China. *Remote Sens.* **2020**, *12*, 236. <https://doi.org/10.3390/rs12020236>.
6. Pollock, R.B.; Kanemasu, E.T. Estimating leaf-area index of wheat with LANDSAT data. *Remote Sens. Environ.* **1979**, *8*, 307–312. [https://doi.org/10.1016/0034-4257\(79\)90030-0](https://doi.org/10.1016/0034-4257(79)90030-0).
7. Trombetta, A.; Iacobellis, V.; Tarantino, E.; Gentile, F. Calibration of the AquaCrop model for winter wheat using MODIS LAI images. *Agric. Water Manag.* **2016**, *164*, 304–316. <https://doi.org/10.1016/j.agwat.2015.10.013>.
8. Boissard, P.; Guérif, M.; Pointel, J.-G.; Guinot, J.-P. Application of SPOT data to wheat yield estimation. *Adv. Sp. Res.* **1989**, *9*, 143–154. [https://doi.org/10.1016/0273-1177\(89\)90479-1](https://doi.org/10.1016/0273-1177(89)90479-1).
9. Becker-Reshef, I.; Vermote, E.; Lindeman, M.; Justice, C. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. *Remote Sens. Environ.* **2010**, *114*, 1312–1323. <https://doi.org/10.1016/j.rse.2010.01.010>.
10. Nasrallah, A.; Baghdadi, N.; El Hajj, M.; Darwish, T.; Belhouchette, H.; Faour, G.; Darwich, S.; Mhawej, M. Sentinel-1 Data for Winter Wheat Phenology Monitoring and Mapping. *Remote Sens.* **2019**, *11*, 2228. <https://doi.org/10.3390/rs11192228>.
11. Hunt, M.L.; Blackburn, G.A.; Carrasco, L.; Redhead, J.W.; Rowland, C.S. High resolution wheat yield mapping using Sentinel-2. *Remote Sens. Environ.* **2019**, *233*, 111410. <https://doi.org/10.1016/j.rse.2019.111410>.
12. Xue, J.; Su, B. Significant remote sensing vegetation indices: A review of developments and applications. *J. Sens.* **2017**, *2017*, 17. <https://doi.org/10.1155/2017/1353691>.
13. Tucker, C.J.; Holben, B.N.; Elgin, J.H.; McMurtrey, J.E. Remote sensing of total dry-matter accumulation in winter wheat. *Remote Sens. Environ.* **1981**, *11*, 171–189. [https://doi.org/10.1016/0034-4257\(81\)90018-3](https://doi.org/10.1016/0034-4257(81)90018-3).
14. Doraiswamy, P.C.; Sinclair, T.R.; Hollinger, S.; Akhmedov, B.; Stern, A.; Prueger, J. Application of MODIS derived parameters for regional crop yield assessment. *Remote Sens. Environ.* **2005**, *97*, 192–202. <https://doi.org/10.1016/j.rse.2005.03.015>.
15. Lopresti, M.F.; Bella, C.M. Di Relationship between MODIS-NDVI data and wheat yield: A case study in Northern Buenos Aires province, Argentina. *Inf. Process. Agric.* **2015**, *2*, 73–84. <https://doi.org/10.1016/j.inpa.2015.06.001>.
16. Asrar, G.; Kanemasu, E.T.; Yoshida, M. Estimates of leaf area index from spectral reflectance of wheat under different cultural practices and solar angle. *Remote Sens. Environ.* **1985**, *17*, 1–11. [https://doi.org/10.1016/0034-4257\(85\)90108-7](https://doi.org/10.1016/0034-4257(85)90108-7).
17. Jin, Z.; Azzari, G.; Lobell, D.B. Improving the accuracy of satellite-based high-resolution yield estimation: A test of multiple scalable approaches. *Agric. For. Meteorol.* **2017**, *247*, 207–220. <https://doi.org/10.1016/j.agrformet.2017.08.001>.
18. Kayad, A.; Sozzi, M.; Gatto, S.; Marinello, F.; Pirotti, F. Monitoring within-field variability of corn yield using sentinel-2 and machine learning techniques. *Remote Sens.* **2019**, *11*, 2873. <https://doi.org/10.3390/rs11232873>.
19. Moriondo, M.; Maselli, F.; Bindi, M. A simple model of regional wheat yield based on NDVI data. *Eur. J. Agron.* **2007**, *26*, 266–274. <https://doi.org/10.1016/j.eja.2006.10.007>.
20. Azzari, G.; Jain, M.; Lobell, D.B. Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries. *Remote Sens. Environ.* **2017**, *202*, 129–141. <https://doi.org/10.1016/j.rse.2017.04.014>.
21. Chlingaryan, A.; Sukkarieh, S.; Whelan, B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Comput. Electron. Agric.* **2018**, *151*, 61–69. <https://doi.org/10.1016/j.compag.2018.05.012>.
22. Filippi, P.; Jones, E.J.; Wimalathunge, N.S.; Somarathna, P.D.S.N.; Pozza, L.E.; Ugbaje, S.U.; Jephcott, T.G.; Paterson, S.E.; Whelan, B.M.; Bishop, T.F.A. An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. *Precis. Agric.* **2019**, *20*, 1015–1029. <https://doi.org/10.1007/s11119-018-09628-4>.
23. Crane-Droesch, A. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environ. Res. Lett.* **2018**, *13*, 114003. <https://doi.org/10.1088/1748-9326/aae159>.
24. Kumhálová, J.; Matějková, Š. Yield variability prediction by remote sensing sensors with different spatial resolution. *Int. Agrophysics* **2017**, *31*, 195–202. <https://doi.org/10.1515/intag-2016-0046>.
25. Paudel, D.; Boogaard, H.; de Wit, A.; Janssen, S.; Osinga, S.; Pylianidis, C.; Athanasiadis, I.N. Machine learning for large-scale crop yield forecasting. *Agric. Syst.* **2021**, *187*, 103016. <https://doi.org/10.1016/j.agry.2020.103016>.
26. Zou, X.; Zhu, S.; Möttus, M. Estimation of Canopy Structure of Field Crops Using Sentinel-2 Bands with Vegetation Indices and Machine Learning Algorithms. *Remote Sens.* **2022**, *14*, 2849. <https://doi.org/10.3390/rs14122849>.
27. Ali, U.; Esau, T.J.; Farooque, A.A.; Zaman, Q.U.; Abbas, F.; Bilodeau, M.F. Limiting the Collection of Ground Truth Data for Land Use and Land Cover Maps with Machine Learning Algorithms. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 333. <https://doi.org/10.3390/ijgi11060333>.
28. Ali, I.; Greifeneder, F.; Stamenkovic, J.; Neumann, M.; Notarnicola, C. Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. *Remote Sens.* **2015**, *7*, 16398–16421. <https://doi.org/10.3390/rs71215841>.
29. Jeong, J.H.; Resop, J.P.; Mueller, N.D.; Fleisher, D.H. Random Forests for Global and Regional Crop Yield Predictions. *PLoS ONE* **2016**, *11*, 1–15. <https://doi.org/10.1371/journal.pone.0156571>.
30. Aha, D.W.; Kibler, D.; Albert, M.K. Instance-based learning algorithms. *Mach. Learn.* **1991**, *6*, 37–66. <https://doi.org/10.1007/BF00153759>.

31. Gonzalez-SanchezAlberto; Frausto-SolisJuan; Ojeda-BustamanteWaldo Predictive ability of machine learning methods for massive crop yield prediction. *Spanish J. Agric. Res.* **2014**, *12*, 313–328. <https://doi.org/10.5424/sjar/2014122-4439>.
32. Segarra, J.; Araus, J.L.; Kefauver, S.C. Farming and Earth Observation: Sentinel-2 data to estimate within-field wheat grain yield. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *107*, 102697. <https://doi.org/10.1016/j.jag.2022.102697>.
33. Van Wittenberghe, S.; Verrelst, J.; Rivera, J.P.; Alonso, L.; Moreno, J.; Samson, R. Gaussian processes retrieval of leaf parameters from a multi-species reflectance, absorbance and fluorescence dataset. *J. Photochem. Photobiol. B Biol.* **2014**, *134*, 37–48. <https://doi.org/10.1016/j.jphotobiol.2014.03.010>.
34. Cavalari, C.; Megoudi, S.; Maxouri, M.; Anatolitis, K.; Sifakis, M.; Levizou, E.; Kyparissis, A. Modeling of durum wheat yield based on sentinel-2 imagery. *Agronomy* **2021**, *11*, 1486. <https://doi.org/10.3390/agronomy11081486>.
35. My John Deere. Available online: <https://myjohndeere.deere.com/> (accessed on 22 May 2022).
36. QGIS.org, 2022. QGIS Geographic Information System. Available online: <http://www.qgis.org> (accessed on 21 May 2022).
37. Open Access Hub. Available online: <https://scihub.copernicus.eu/> (accessed on 22 May 2022).
38. STEP—Science Toolbox Exploitation Platform. Available online: <http://step.esa.int> (accessed on 22 May 2022).
39. Jia, P.; Zhang, J.; He, W.; Hu, Y.; Zeng, R.; Zamanian, K.; Jia, K.; Zhao, X. Combination of Hyperspectral and Machine Learning to Invert Soil Electrical Conductivity. *Remote Sens.* **2022**, *14*, 2602. <https://doi.org/10.3390/rs14112602>.
40. JASP—A Fresh Way to Do Statistics le. Available online: <https://jasp-stats.org> (accessed on 22 May 2022).
41. Appelhans, T.; Mwangomo, E.; Hardy, D.R.; Hemp, A.; Nauss, T. Evaluating machine learning approaches for the interpolation of monthly air temperature at. *Spat. Stat.* **2015**, *14*, 91–113. <https://doi.org/10.1016/j.spasta.2015.05.008>.
42. Chen, S.; Liu, W.; Feng, P.; Ye, T.; Ma, Y.; Zhang, Z. Improving Spatial Disaggregation of Crop Yield by Incorporating Machine Learning with Multisource Data: A Case Study of Chinese Maize Yield. *Remote Sens.* **2022**, *14*, 2340. <https://doi.org/10.3390/rs14102340>.
43. Basso, B.; Dumont, B.; Cammarano, D.; Pezzuolo, A.; Marinello, F.; Sartori, L. Environmental and economic benefits of variable rate nitrogen fertilization in a nitrate vulnerable zone. *Sci. Total Environ.* **2016**, *545–546*, 227–235. <https://doi.org/10.1016/j.scitotenv.2015.12.104>.
44. Guerrero, A.; Mouazen, A.M. Evaluation of variable rate nitrogen fertilization scenarios in cereal crops from economic, environmental and technical perspective. *Soil Tillage Res.* **2021**, *213*, 105110. <https://doi.org/10.1016/j.still.2021.105110>.
45. Stettmer, M.; Maidl, F.-X.; Schwarzensteiner, J.; Hülsbergen, K.-J.; Bernhardt, H. Analysis of Nitrogen Uptake in Winter Wheat Using Sensor and Satellite Data for Site-Specific Fertilization. *Agronomy* **2022**, *12*, 1455. <https://doi.org/10.3390/agronomy12061455>.
46. Uribeetxebarria, A.; Castellón, A.; Aizpurua, A. A First Approach to Determine If It Is Possible to Delineate In-Season N Fertilization Maps for Wheat Using NDVI Derived from Sentinel-2. *Remote Sens.* **2022**, *14*, 2872. <https://doi.org/10.3390/rs14122872>.
47. Waldamichael, F.G.; Debelee, T.G.; Schwenker, F.; Ayano, Y.M.; Kebede, S.R. Machine Learning in Cereal Crops Disease Detection: A Review. *Algorithms* **2022**, *15*, 1–19. <https://doi.org/10.3390/a15030075>.
48. Cravero, A.; Pardo, S.; Sepúlveda, S.; Muñoz, L. Challenges to Use Machine Learning in Agricultural Big Data: A Systematic Literature Review. *Agronomy* **2022**, *12*, 748. <https://doi.org/10.3390/agronomy12030748>.