Combining PlanetScope and Sentinel-2 images with environmental data for improved wheat yield estimation

Nizom Farmonov [a,b], Khilola Amankulova [a], József Szatmári [a], Jamol Urinov[c], Zafar Narmanov[d], Jakhongir Nosirov[c] and László Mucsi [a]

aDepartment of Geoinformatics, Physical and Environmental Geography, University of Szeged, Szeged, Hungary; bTashkent Institute of Irrigation and Agricultural Mechanization Engineers, National Research University, Tashkent, Uzbekistan; cDepartment of Land resources, Cadastre and Geoinformatics, Karshi Institute of Irrigation and Agrotechnology "TIIAME" National Research University, Karshi, Uzbekistan; dSoil Science Department of the Kashkadarya Region regional division of the Institute of Soil Science and Agrochemical Research, Karshi, Uzbekistan

ABSTRACT

Satellite images are widely used for crop yield estimation, but their coarse spatial resolution means that they often fail to provide detailed information at the field scale. Recently, a new generation of high-resolution satellites and CubeSat platforms has been launched. In this study, satellite data sources including PlanetScope and Sentinel-2 were combined with topographic and climatic variables, and the improvement in wheat yield estimation was evaluated. Wheat yield data from a combine harvester were used to train and validate a yield estimation model based on random forest regression. Nine vegetation indices (NDVI, GNDVI, MSAVI2, MTVI2, MTCI, reNDVI, SAVI, EVI and WDVI) and spectral bands were tested. During the model training, the Sentinel-2 data realized a slightly higher estimation accuracy than the PlanetScope data. However, combining environmental data with the PlanetScope data realized the highest estimation accuracy. For the validated models, adding the topographic and climatic datasets to the satellite data sources improved the estimation accuracy, and the results were slightly better with the Sentinel-2 data than with the PlanetScope data. Observation data of the milk and dough stages provided the highest estimation accuracy of the wheat yield at 210–225 days after sowing.

1. Introduction

Wheat is a major food source and the world's fifth-largest crop by consumption (Igrejas and Branlard 2020). In 2017, wheat was cultivated over an area of more than 220 million ha, and 757 million tons was produced, which is third after maize and rice among cereals. Early prediction of the crop yield is becoming increasingly important owing to the inconsistent conditions around the world to ensure food security and farmers' revenue (Rosegrant and Cline 2003; Ittersum 2016; Balogh and Novák 2020). The creation of site-specific management strategies can be aided by the identification of yield-limiting variables for reliable yield estimation. Thus, obtaining a high-

CONTACT Nizom Farmonov ✉ farmonov.nizom@stud.u-szeged.hu Department of Geoinformatics, Physical and Environmental Geography, University of Szeged, Egyetem utca 2, 6722 Szeged, Hungary; Tashkent Institute of Irrigation and Agricultural Mechanization Engineers, National Research University, 39 Kari Niyazov Street, Tashkent 100000, Uzbekistan

resolution estimation of the crop yield prior to the harvest is a key crop management tool to ensure agricultural, environmental, and socioeconomic sustainability.

Earth observation (EO) refers to satellite-based technologies that are used to collect information on Earth. Such technologies can be used to track crop development and yields, which can help guide the development of site-specific management techniques for smart farming (Hunt et al. 2019). Several EO systems have been developed to monitor crops and estimate crop yields, such as the Advanced Very High-Resolution Radiometer (AVHRR) and Moderate Resolution Imaging Spectroradiometer. However, the coarse spatial resolution means that such data often fail to provide details at the field scale.

Recently, satellites have been launched that can facilitate improved estimation accuracy of the crop yield. Various optical sensors have been explored for their potential application to estimating crop yields. Sentinel-2 was launched in 2015 by the European Space Agency, and the spatial and spectral resolutions of its images support the monitoring of vegetation in various growth stages and provide new opportunities for yield estimation and crop monitoring. In addition, CubeSat constellations such as the commercially available PlanetScope now offer data with a high combination of temporal and spatial resolutions.

Several studies have considered using high-resolution satellite imagery to estimate the wheat yield at the field scale. Hunt et al. (2019) used Sentinel-2 and random forest regression (RFR) to successfully estimate the wheat yield in the United Kingdom at a resolution of 10 m. They found that images obtained in June provided an accurate estimation of the yield and that collecting Sentinel-2 data during the growing season improved the estimation accuracy. Their results indicated that the estimation accuracy could be further improved by combining the satellite imagery with environmental data. Zhao et al. (2020) combined Sentinel-2 data and a crop model to estimate the wheat yield in northeastern Australia. They considered potential vegetation indices (VIs) related to the canopy structure and chlorophyl derived from the Sentinel-2 images. Their results showed that the red edge chlorophyl index (CI) ($R^2 = 0.76$, root mean square error (RMSE) = 0.88 t/ha) and optimized soil-adjusted vegetation index (OSAVI) ($R^2 = 0.74$, RMSE = 0.91 t/ha) provided the best estimation accuracy. In addition, the estimation accuracy increased when both structural and chlorophyl indices were combined ($R^2 = 0.91$ and RMSE = 0.54 t/ha).

Jeong et al. (2016) evaluated the effectiveness of RFR at estimating the yield of staple crops. Among various machine-learning (ML) techniques, random forest has been shown to be well suited to estimating crop yields (Rehfeldt et al. 2012; Singh, Sihag, and Singh 2017; Han et al. 2020; Maya Gopal and Bhargavi 2019). Despite the availability of Sentinel-2 and CubeSat data, their efficiency has not fully explored and compared, and further research is necessary. Only a few studies have focused on estimating the wheat yield based on satellite data and measurements from global positioning system (GPS) combine harvesters. Manivasagam et al. (2021) combined Sentinel-2 and PlanetScope image data with the Simple Algorithm for Yield Estimate Crop Model, which they applied to estimating the wheat yield in Israel. Their results indicated that the leaf area index (LAI) derived from both PlanetScope and Sentinel-2 data provided a higher yield estimation accuracy (RMSE = 69 g/m^2) than that of LAI derived from Sentinel-2 alone (RMSE = 88 g/m^2). However, no studies have directly assessed the potential of Sentinel-2 and PlanetScope image data at estimating the wheat yield at the field scale with combine harvesters. This paper demonstrates a new method of integrating wheat grain yield, global positioning systems, combine harvester data, climate, topographic variables, and high spatial–temporal resolution of satellite imagery.

In this study, we evaluated the effectiveness of PlanetScope and Sentinel-2 image data at wheat yield estimation and the influence of environmental variables on the estimation accuracy. The next-generation instrument PlanetScope SuperDove (PSB.SD) with eight spectral bands was used for the first time for wheat yield estimation. Seven wheat fields were monitored with PlanetScope and Sentinel-2, and four VIs were derived to describe the phenological stages in 2020–2021. Different combinations of the satellite data, VIs, and environmental data (e.g. topography and climate) were

applied to an RFR model, and the yield estimation accuracy was evaluated. The study was designed to answer four key questions:

(1) How do the spatial and temporal resolutions of PlanetScope and Sentinel-2 affect the yield estimation accuracy?
(2) Does the addition of VIs contribute extra information to the estimation model?
(3) How does combining the PlanetScope and Sentinel-2 data with environmental data affect the estimation accuracy?
(4) Which phenological stage and VI provide an accurate estimation of the wheat yield?

2. Materials and methods

2.1. Study area

The fields considered in this study are located in Mezhegyes experimental farm, Mezhegyes Town, Békés County, of southeast Hungary (46°19′ N, 20°49′ E), which is near the Romanian border (Figure 1). The town has a population of 4950 people and a total administrative area of 15544 ha. Four fields were chosen for analysis. Two fields were used to train the model, and the other two were used for validation. Wheat is the most widely grown crop and accounts for 1223 ha of the total area of the farm. The average field size is 30 ha with a maximum of 100 ha. The main soil type is chernozem, which is suitable for growing crops owing to the high levels of lime particularly cereals and oilseeds. In 2021, the average annual rainfall was 645 mm (428.9 mm in-crop). The average annual temperature in the study area ranges from 7.8°C to 11.1°C.

2.2. Field data

The ground truth data on the wheat yield were obtained between June 29 and July 24, 2021, by using a John Deere W650i combine harvester equipped with a GPS and yield monitoring system. Wheat
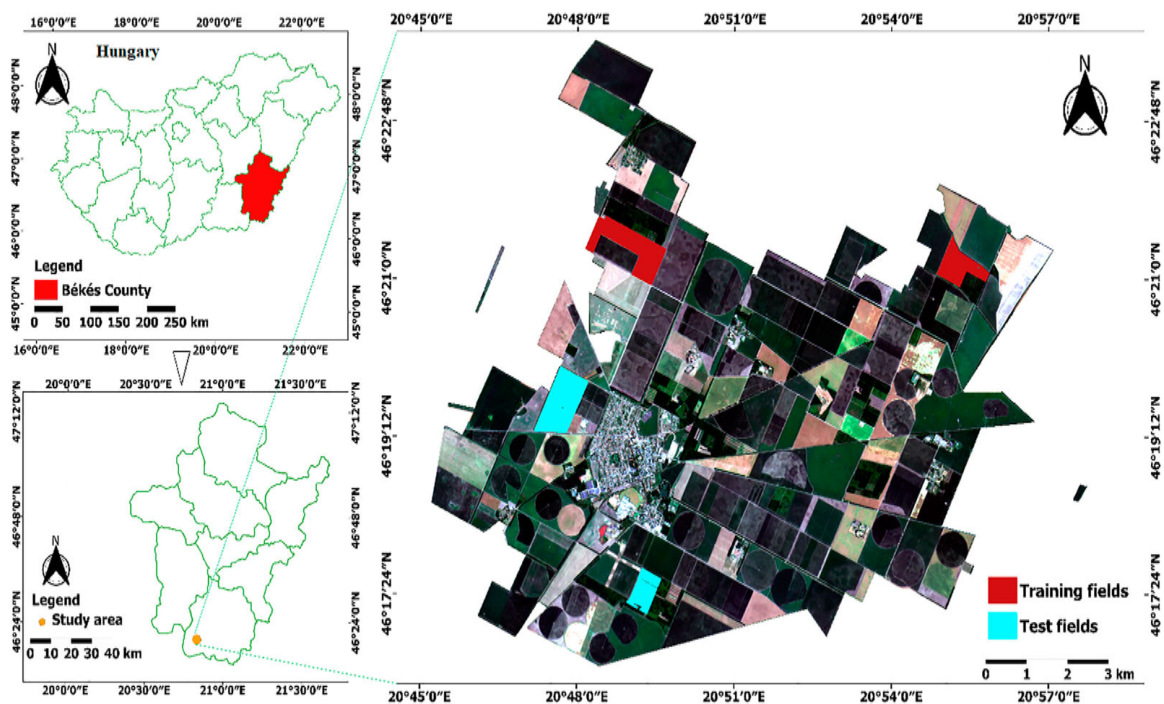
Figure 1. Study site. The RGB true color composite image was acquired by PlanetScope on June 28, 2021.

seeds were sown in November and harvested in July. The average wheat yield was ∼4.5–5 t/ha, and the maximum yield was ∼10 t/ha. After the monitoring data were collected, they were filtered, edited, and converted to yield maps. The process of cleaning inaccurate grain yield includes determining combine delay times and removing 'overlapping' data, especially near the end of rows. In QGIS, all GPS crop yield points obtained from the combine harvester were uploaded as shapefiles. As the information was organized in attribute tables, it was easier to process and filter. First, zero and near-zero yield points were removed from the attribute table. Second, using the combine harvester's header, we selected homogeneous yield points with the same distance and swath width. Subsequently, the edge of the fields was buffered to avoid mixed pixels. The raw yield maps were exported as shapefiles (.shp), which can be manipulated in a geographic information system. Commercial yield monitoring systems are prone to enormous errors typically caused by the incorrect calibration of multiple machines, incorrect setting of the header height and cut width, and incorrect adjustment of the speed and travel distance (Kharel et al. 2019; Matcham et al. 2022). Data points that did not have x and y GPS coordinates were deleted. Finally, the point vector layer was converted in QGIS to a raster grid by using inverse distance weighted interpolation to pixel sizes of 3 and 10 m to match the spatial resolutions of PlanetScope and Sentinel-2, respectively.

2.3. Remote sensing data

2.3.1. Planetscope

We downloaded 72 cloud-free PlanetScope (PS) level-3 surface reflectance products collected between November 2020 and July 2021 from the Planet Explorer website (https://www.planet.com/explorer/ accessed on August 25, 2022). The spectral data were obtained from PSB.SD, which has eight spectral bands (i.e. red edge, red, green, green I, yellow, blue, coastal blue, and near-infrared) with a pixel size of 3 m and near-daily global time revisit. The orthorectified product was corrected in terms of geometry, radiometry, and cartography for surface reflection and was projected to a UTM/WGS84 cartographic map (Planet Team 2017). The images were harmonized with Sentinel-2 to ensure consistent radiometry. We performed it during the ordering of PS scenes through the 'harmonize' imagery option available in Planet Explorer. There are three generations of Doves operated by Planet (Dove Classics, Dove-R, SuperDove) with varying spectral responses. Using the 'harmonize' option provides data that is approximately the same spectral response as Sentinel-2. Harmonization is applied to all Doves and allows all sensors to be consistent and compatible with each other. The first coastal blue band was eliminated. Finally, the VIs and phenological stages were derived by stacking the PS bands together as layers. Table 1 lists the PS data used in this study.

2.3.2. Sentinel-2

We downloaded 25 cloud-free Sentinel-2 (S2) level-2A satellite images from the Copernicus Open Access Hub website (https://scihub.copernicus.eu/dhus/#/home accessed on 5 September 2022) covering the study period. A level-2A product provides images of the bottom of atmosphere reflectance and covers the visible and near-infrared spectral ranges derived from associated level-1C datasets. S2 A and B are equipped with multispectral instruments that provide advanced crop

Table 1. Description of PS data used in this study.

Bands	Central Wavelength (nm)	Spatial Resolution (m)
Band 2 - Blue	490 (50)	3
Band 3 - Green I	531 (36)	3
Band 4 - Green	565 (36)	3
Band 5 - Yellow	610 (20)	3
Band 6 - Red	665 (31)	3
Band 7 - Red Edge	705 (15)	3
Band 8 - NIR	865 (40)	3

monitoring at the field, regional, and global scales with various spatial resolutions (10, 20, and 60 m) (Vijayasekaran 2019). Bands with resolutions of 20 and 60 m were resampled to 10 m by using nearest-neighbor interpolation in the Sentinel Applications Platform (SNAP) version 8.0 (https://step.esa.int accessed on 1 September, 2021) to ensure that all channels were concatenated with aligned pixels. All bands were stacked as layers into one file to produce a time series of S2 images. Table 2 explains the S2 bands that were used in this study.

2.4. Vegetation indices

Table 3 lists the nine VIs used in this study, which has been used in previous studies to estimate the wheat yield (Han et al. 2020; Segarra, Luis Araus, and Kefauver 2022; Zhao et al. 2020). The normalized difference vegetation index (NDVI) (Tucker 1979) and green normalized difference vegetation index (GNDVI) (Gitelson, Kaufman, and Merzlyak 1996) are widely used to determine the amounts of water and nitrogen taken up by crops (Tucker 1979; Shanahan et al. 2001; Jackson 2004). Gitelson, Kaufman, and Merzlyak (1996) developed the GNDVI to address saturation issues observed with NDVI for some vegetation types in later growth stages. Soil adjusted vegetation index (SAVI) and the modified soil-adjusted vegetation index (MSAVI2) is used in the early stages of crop growth to track emerging seedlings. It reduces the impact of bare soil, so it is ideal for early growth stages such as crop emergence, for monitoring crops that do not cover the soil even in the most advanced growth stage, and for woody crops (Qi et al. 1994). The modified triangular vegetation index (MTVI2) is nearly equivalent to MTVI but is considered a better estimator of green LAI. It accounts for the soil background signature while retaining sensitivity to LAI and robustness against the influence of chlorophyl (Haboudane 2004). The MERIS terrestrial chlorophyl index (MTCI) can be used in precision agriculture because it is more sensitive to high chlorophyl content (Dash and Curran 2007). Red edge normalized difference vegetation index (reNDVI) measures the sensitivity of canopy foliage, gap fraction, and senescence to altering its red edge (Gitelson and Merzlyak 1994). Enhanced vegetation index (EVI) is effective to mitigate saturation because it takes into account some atmospheric conditions and canopy background noise, particularly in areas with dense vegetation. Weighted difference vegetation index (WDVI) is suitable for the correction of soil background and is very sensitive to atmospheric variations.

The crop phenology is dynamic during the growing season (Ruml and Vulic 2005). Twice a month, the four wheat fields were observed for changes in phenology, and the transition dates were recorded. The field measurements and satellite-derived spectral reflectance patterns were compared. NDVI, GNDVI, MSAVI2, MTVI2, MTCI, reNDVI, SAVI, EVI and WDVI and were analyzed to identify phenological patterns. MTVI2 was used to measure and assess the leaf chlorophyll content at the canopy scale owing to its insensitivity to the LAI. All satellite images were used to extract time series of the VIs. The BBCH-scale was used to distinguish the phenological development and growth stages of the wheat, as described in Table 6 (Lancashire et al. 1991). Points were obtained by using the random points inside the polygon tool in QGIS 3.16. The values

Table 2. Description of S2 bands used in this study.

Bands	Central Wavelength (nm)	Spatial Resolution (m)
Band 2 - Blue	490 nm	10 m
Band 3 - Green	560 nm	10 m
Band 4 - Red	665 nm	10 m
Band 5 - Vegetation red edge	705 nm	20 m
Band 6 - Vegetation red edge	740 nm	20 m
Band 7 - Vegetation red edge	783 nm	20 m
Band 8 - NIR	842 nm	10 m
Band 8A - Narrow NIR	865 nm	20 m
Band 11 - SWIR	1610 nm	20 m
Band 12 - SWIR	2190 nm	20 m

Table 3. Multispectral VIs were used in this study.

Index	Equation	References
Normalized difference vegetation index (NDVI)	$\dfrac{NIR - Red}{NIR + Red}$	(Rouse et al. 1973)
Green normalized difference vegetation index (GNDVI)	$\dfrac{NIR - Green}{NIR + Green}$	(Gitelson, Kaufman, and Merzlyak 1996)
Modified soil-adjusted vegetation index (MSAVI2)	$(NIR + 1) - \left(\dfrac{1}{2}\right)[(2NIR + 1) - 8(NIR - R)]1/2$	(Qi et al. 1994).
Modified triangular vegetation index (MTVI2)	$\dfrac{1.5[1.2(NIR - Green) - 2.5(Red - Green)]}{\sqrt{(2NIR + 1)^2 - (6NIR - 5\sqrt{Red})} - 0.5}$	(Haboudane 2004).
MERIS terrestrial chlorophyl index (MTCI)	$\dfrac{NIR - RE}{RE - RED}$	(Dash and Curran 2007)
Red edge normalized difference vegetation index (reNDVI)	$\dfrac{NIR - RE}{NIR + RE}$	(Gitelson and Merzlyak 1994)
Soil adjusted vegetation index (SAVI)	$\dfrac{(NIR - RED)}{(NIR + RED + L)} * (1 + L)$	(Huete 1988)
Enhanced vegetation index (EVI)	$\dfrac{2.5 * (NIR - RED)}{(NIR + 6 * RED - 7 * BLUE + 1)}$	(Huete et al. 2002)
Weighted difference vegetation index (WDVI)	$NIR - S * RED$ where: S is the slope of the soil line	(Clevers 1989)

of the VIs were extracted from the wheat fields by using the point sampling tool and open-source plugin in QGIS to determine the crop phenology and transition dates. The 85 points that were generated randomly in QGIS for each multi-temporal VIs inside wheat field boundaries were then averaged and distributed over the wheat growth stages to create phenological phases. The crop ages in the satellite images were calculated according to days after sowing (DAS).

2.5. Environmental data

2.5.1. Climatic variables

Monthly (1/24°, 4 km) gridded TerraClimate datasets for the total precipitation (mm) and maximum temperature (°C) were downloaded from the Google Earth Engine cloud platform (Abatzoglou et al. 2018). TerraClimate incorporates a monthly climate and climatic water balance covering global terrestrial surfaces from the University of California Merced as well as various high- and coarse-resolution climatological datasets (e.g. WorldClim, Japanese 55-year Reanalysis (JRA55)). Monthly data for June 2021 were obtained that had a relatively high spatial resolution when compared to other climatic datasets. The high spatial distribution allowed us to detect spatial variations in the rainfall and temperature across the study area. The datasets were fed into the estimation model as an input feature.

2.5.2. Topographic variables

A high-accuracy LiDAR digital terrain model (DTM) with a spatial resolution of 5 cm was obtained for the study area. The DTM was derived from airborne radar data collected in 2019. The data were resampled by using the cubic convolution method in the software ERDAS IMAGINE 2020 to resolutions of 3 and 10 m to match the spatial resolutions of PS and S2, respectively. The rescaled datasets were used to calculate the secondary variables, slope, and aspect as input parameters for the estimation model.

2.6. Estimation model based on RFR

RFR is an ensemble learning method that is based on the decision tree algorithm. It can be used for both classification and regression tasks, and it has been applied to crop yield estimation (Smith, Ganesh, and Liu 2013). In RFR, the model builds tree predictors associated with different random vector values sampled independently. The model constructs decor-related decision trees during the training phase, and the overall model output is obtained by averaging the output values of all

Table 4. Independent variables were used to build the RFR model.

Variable type		Dataset	Pixel size	Temporal resolution
PlanetScope		Dove PSB.SD bands: 2, 3, 4, 5, 6, 7, 8	3m	November 6, 2020 - July 31, 2021
Sentinel-2		Sentinel-2 Level 2A bands: 2, 3, 4, 5, 6, 7, 8, 8a, 11, 12	10 m	November 6, 2020 – July 30, 2021
Vegetation indices		NDVI, GNDVI, MSAVI2, MTVI2, MTCI, reNDVI, SAVI, EVI and WDVI derived from PlanetScope and Sentinel-2 images	3, 10 m	November 2020 – July 2021
Environmental	Precipitation	TerraClimate: monthly accumulated precipitation (mm)	~4km	November 2020 – July 2021
	Temperature	TerraClimate: monthly accumulated maximum temperature (°C)		
	DTM	Airborne LiDAR	5cm (resampled to 3 and 10 m)	Created using the data acquired in 2019
	Aspect	Calculated using the LiDAR DTM		
	Slope	Calculated using the LiDAR DTM		

individual trees. The learner bagging algorithm is used to train any single tree (Breiman 2001). RFR combines the estimations from multiple ML algorithms to improve the accuracy over that of a single model, which is the main benefit compared with decision trees (Fawagreh, Gaber, and Elyan 2014).

The RFR model was implemented by using the randomForest package in R software (Liaw and Wiener 2002). In this study, the default parameters of the random forest ntree and mtry were used for the regression analysis. The number of trees produced in the regression forest (i.e. ntree) was set at 500, and the number of distinct predictors sampled at each node (i.e. mtry) was set to a default of the number of predictors (i.e. 203) divided by 3. Firstly, four wheat fields were selected for analysis which has GPS combine harvester yield data. We used two merged wheat parcels for model development by dividing 70% for training and 30% for the test, respectively. Furthermore, owing the best-performed regression models were applied for validation parcels to verify the result that has completely independent datasets which included the two fields. Several combinations of variables were applied to the RFR model to determine the best estimation accuracy at the field level. Table 4 presents the datasets used as independent variables during the model training to estimate field-scale variability of the wheat yield. Table 5 presents the layer combinations used to examine how different combinations of data and temporal coverages affected the estimation accuracy.

The yields estimated from the training and validation datasets were compared to the observed yields collected by the combine harvester, and the residuals were calculated. We calculated the

Table 5. Data combinations applied to the RFR model.

Integration	Data layers
Question 1: Sensor comparison	
PS	PlanetScope bands
S2	Sentinel-2 bands
Question 2: Testing VIs individually and in combination with spectral bands of PS and S2	
VI	VIs extracted from PS and S2
PS-VI	PlanetScope + VIs
S2-VI	Sentinel-2 + VIs
Question 3 Combination of the Topographic and climate data to the best-performed integrated Spectral bands and VIs	
PS-VI – Topographic	PlanetScope + VIs + DTM, Aspect, Slope
PS-VI - Topographic-Climate	PlanetScope + VIs + DTM + Aspect + Slope + Precipitation + Temperature
S2-VI – Topographic	Sentinel-2 + VIs + DTM, Aspect, Slope
S2-VI - Topographic-Climate	Sentinel-2 + VIs + DTM + Aspect + Slope + Precipitation + Temperature
Question 4: Identification of best-performed single VIs and growing stage	
PS	NDVI, GNDVI, MSAVI2, MTVI2, MTCI, reNDVI, SAVI, EVI and WDVI
S2	NDVI, GNDVI, MSAVI2, MTVI2, MTCI, reNDVI, SAVI, EVI and WDVI

Table 6. Phenological development of wheat from sowing to harvest.

Month	Description
01–15 November	Emergence
01–31 December	Tillering
01-January 29-February	Vernalization & Tillering
01–31 March	Tillering
01–31 April	Steam Elongation
01–31 May	Botting & Flowering & Fertilization
01–31 June	Milk & Dough stage
July	Physiological Maturity & Harvest

coefficient of determination (R^2) and RMSE as evaluation metrics for the estimation accuracy of the model:

$$R^2 = 1 - \frac{RSS}{TSS}. \tag{1}$$

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

3. Results

3.1. Phenology monitoring

The four VIs were used to describe the wheat growth stages from sowing to harvest. Identifying the growth stage that most affects the potential yield can help improve the effectiveness of management actions. For instance, the yield is affected depending on which growth stage is subjected to actions such as fertilization or pesticides or stresses such as frost, hail, low soil moisture, and plant diseases. Figure 2 illustrates the different temporal patterns acquired from the satellite data. The VIs derived from PS and S2 data demonstrated nearly identical and consistent temporal patterns. All VIs showed their lowest values at the beginning of the vegetative period. The VIs began to steadily increase after a few weeks (30–73 DAS), which indicated the initiation of vegetative stages (e.g. leaf development) and significant growth. The wheat growth peaked at 155–230 DAS, which corresponded to the highest VI values. The wheat had begun to fruit when the VIs started to decline at 231–243 DAS. The VIs then dropped to their lowest values at 244–275 DAS, which corresponded to the when the wheat had ripened and was harvested.

3.2. Model training

PS and S2 data obtained for each phenological stage were analyzed individually to identify any strong correlations with the wheat yield. The estimation accuracy of the RFR model showed the highest correlation with the yield data when using spectral bands from the botting, flowering, and fertilization

Table 7. RFR model performance at each phenological stage using PS bands.

BBCH-scale	Day	RMSE	R2
Emergence	11-November 2020	0.514	0.44
Tillering	5-December 2020	0.574	0.32
Vernalization	27-January	0.591	0.28
Tillering	15-February	0.508	0.46
Tillering	21-March	0.523	0.43
Steam Elongation	10-April	0.516	0.44
Botting & Flowering & Fertilization	**27-May**	**0.380**	**0.70**
Milk & Dough stage	**17-June**	**0.336**	**0.79**
Physiological Maturity & Harvest	25-July	0.459	0.56
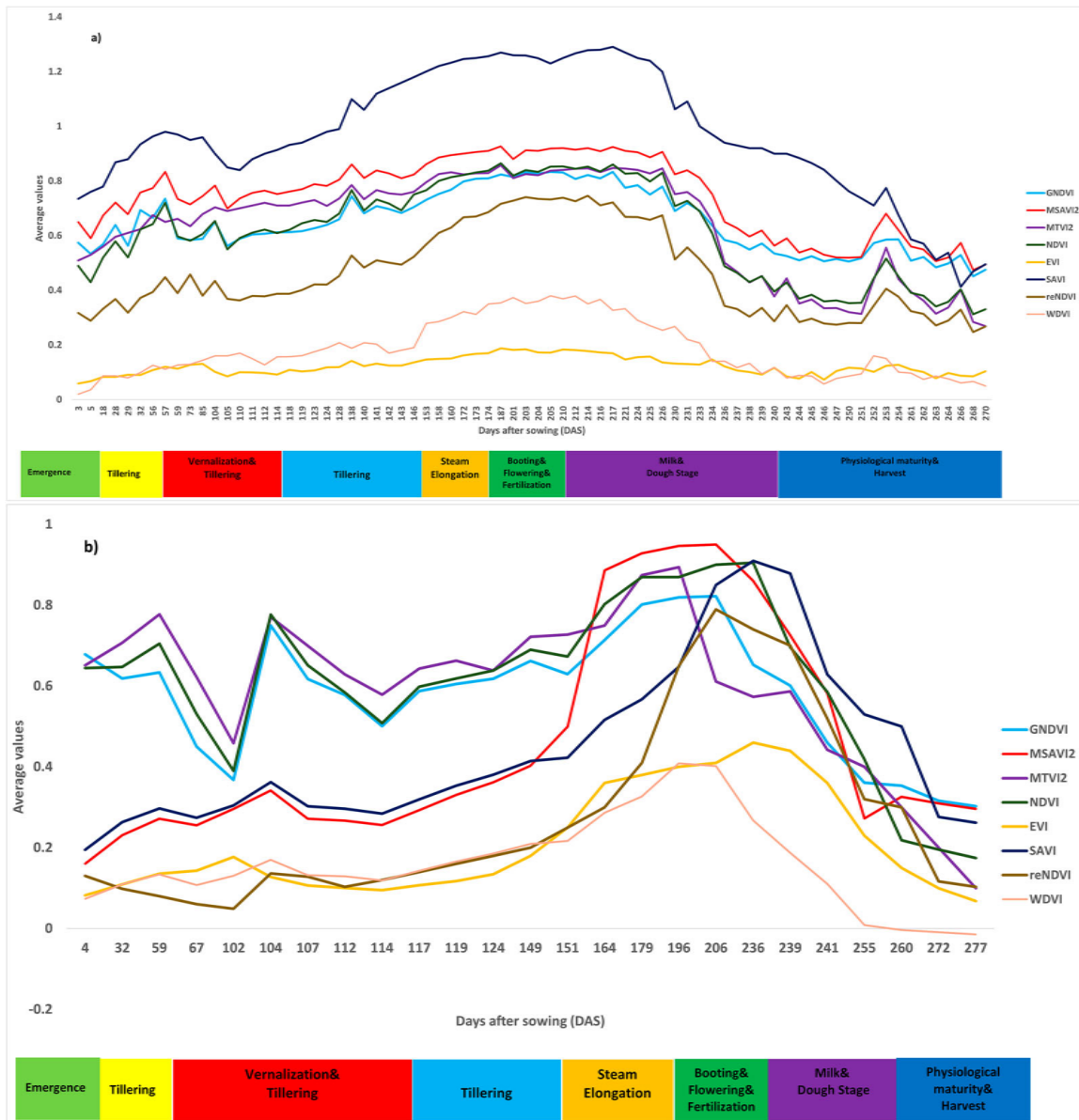
Figure 2. Wheat growth at the pixel level according to the BBCH-scale: (a) PS and (b) S2. For each wheat yield pixel, four VIs were extracted from sowing to harvest. The horizontal axis indicates DAS.

stage and the milk and dough stages, with RMSE values of 0.336–0.380 t/ha with PS data and 0.325–0.389 t/ha with S2 data. As indicated in Tables 7 and 8, the estimation accuracy of the model continued to increase until July, which corresponded to the peak growth of the wheat. However, the estimation accuracy did not increase any further when data from July onward were incorporated.

Table 8. RFR model performance at each phenological stage using S2 bands.

BBCH-scale	Day	RMSE	R2
Emergence	7-November 2020	0.391	0.70
Tillering	5-December 2020	0.424	0.65
Vernalization	9-January	0.377	0.70
Tillering	15-February	0.394	0.69
Tillering	7-March	0.361	0.72
Steam Elongation	9-April	0.377	0.71
Botting & Flowering & Fertilization	**21-May**	**0.389**	**0.79**
Milk & Dough stage	**20-June**	**0.325**	**0.81**
Physiological Maturity & Harvest	25-July	0.375	0.71

Table 10. RMSE and R-squared values of the RFR model using S2 bands and VIs at the peak phenological stage.

Sentinel 2	Band		Indices	
Days	RMSE	R^2	RMSE	R^2
11-may	0.373	0.78	0.468	0.57
21-may	0.389	0.79	0.452	0.59
20-june	**0.325**	**0.81**	**0.426**	**0.66**
23-june	0.348	0.80	0.453	0.58
25-june	0.354	0.80	0.459	0.60

show scatter plots demonstrating the relationship between the observed and estimated wheat yields of individual validation fields according to the differing satellites. The estimated wheat distribution visually reflected the general patterns of the observed yield with relatively little field-scale variation. We were also able to identify areas where the model underestimated and overestimated yields. Overall, the model estimated the field-scale yield variability with reasonable accuracy, and the RMSE values differed between fields.

The results can be summarized by answering the four key questions around which the study was based as follows:

1. How do the spatial and temporal resolutions of PS and S2 affect the yield estimation accuracy?

The training results showed that high-resolution PS and S2 image data (3 and 10 m, respectively) could be used to map the wheat yield with similar results (R^2 = 0.79 and 0.81, respectively) (Tables 9 and 10).

2. Does the addition of VIs contribute extra information to the estimation model?

The training results showed that the RMSE value was lower with the spectral bands than with the VIs, and the error was slightly higher for the VIs alone. When the spectral bands and VIs were combined, the yield estimation accuracy increased marginally with the PS-based model but not always with the S2-based model for both training and validation (Table 11,). This indicates that adding VIs to the spectral bands added some insights that helped improve the estimation accuracy of the model.

3. How does combining the PS and S2 data with environmental data affect the estimation accuracy?

The estimation accuracy of the model was increased by combining the environmental data with the spectral bands and VIs (Table 11). Adding topographic variables such as the DTM and the slope and aspect noticeably increased the model accuracy. Further improvement was achieved by adding climatic data to the model (e.g. monthly precipitation and temperature).

4. Which phenological stage and VI provide an accurate estimation of the wheat yield?

Table 11. Combinations of the spectral band, VI, and environmental data for the optimal dates used to train the RFR models.

|	PlanetScope		Sentinel 2	
Data layers	RMSE	R^2	RMSE	R^2
Bands	0.336	0.79	0.325	0.81
Vis	0.412	0.64	0.426	0.66
Bands + VIs	0.321	0.81	0.321	0.80
Bands + VIs + Topographic	0.307	0.83	0.309	0.83
Band + VIs + Topographic + Climate	0.287	0.84	0.298	0.84
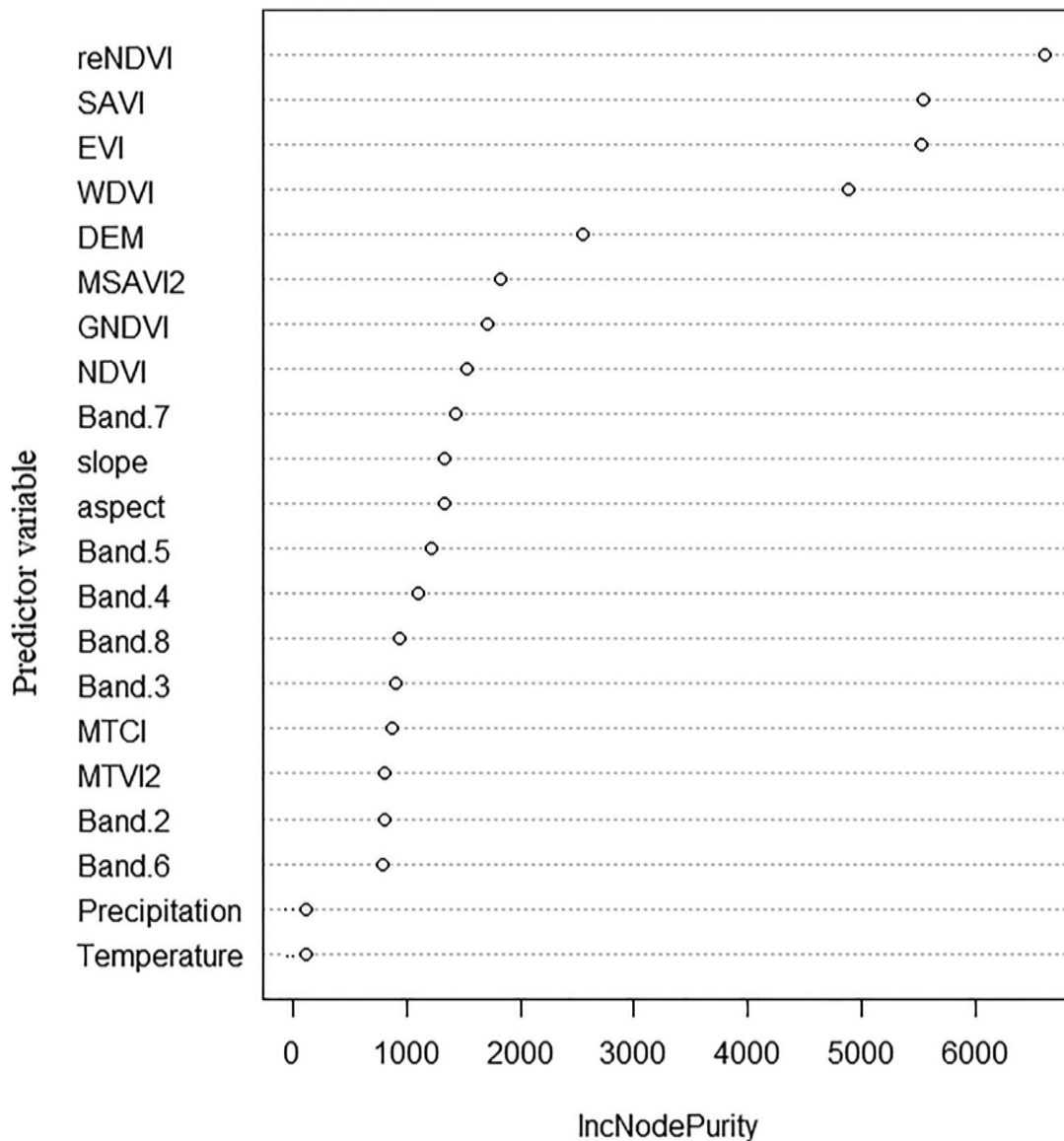
Figure 3. Variable importance list of all RFR model (PS + VIs + Topographic + Climate).

Data from the milk and dough stage (210–225 DAS, BBCH-scale: 71–83) provided higher estimation accuracy of the wheat yield. Among nine VIs, MTCI derived from PS performed better in yield estimation with RMSE = 0.500 t/ha, followed by MTVI, SAVI and reNDVI, respectively (Figure 9). In the case of S2 VIs, MTVI2 produced high accuracy with RMSE = 0.550 t/ha, followed by reNDVI, GNDVI and MTCI.

4. Discussion

4.1. Advantages of RFR compared with other ML methods

This study investigated the suitability of PS and S2 image data for estimating the wheat yield at the field scale. We selected the RFR model for the yield estimation because we discovered that the correlation between the crop yield and reflectance is sufficiently sophisticated for ML methods. Because RFR is less likely to contain outliers, we expected it to realize a higher estimation accuracy than other ML methods. RFR can be used to effectively handle both linear and nonlinear relationships as well as insensitivity to overfitting where linear regression fails (Kumhálová and Matějková 2017; Xie et al. 2021). It is
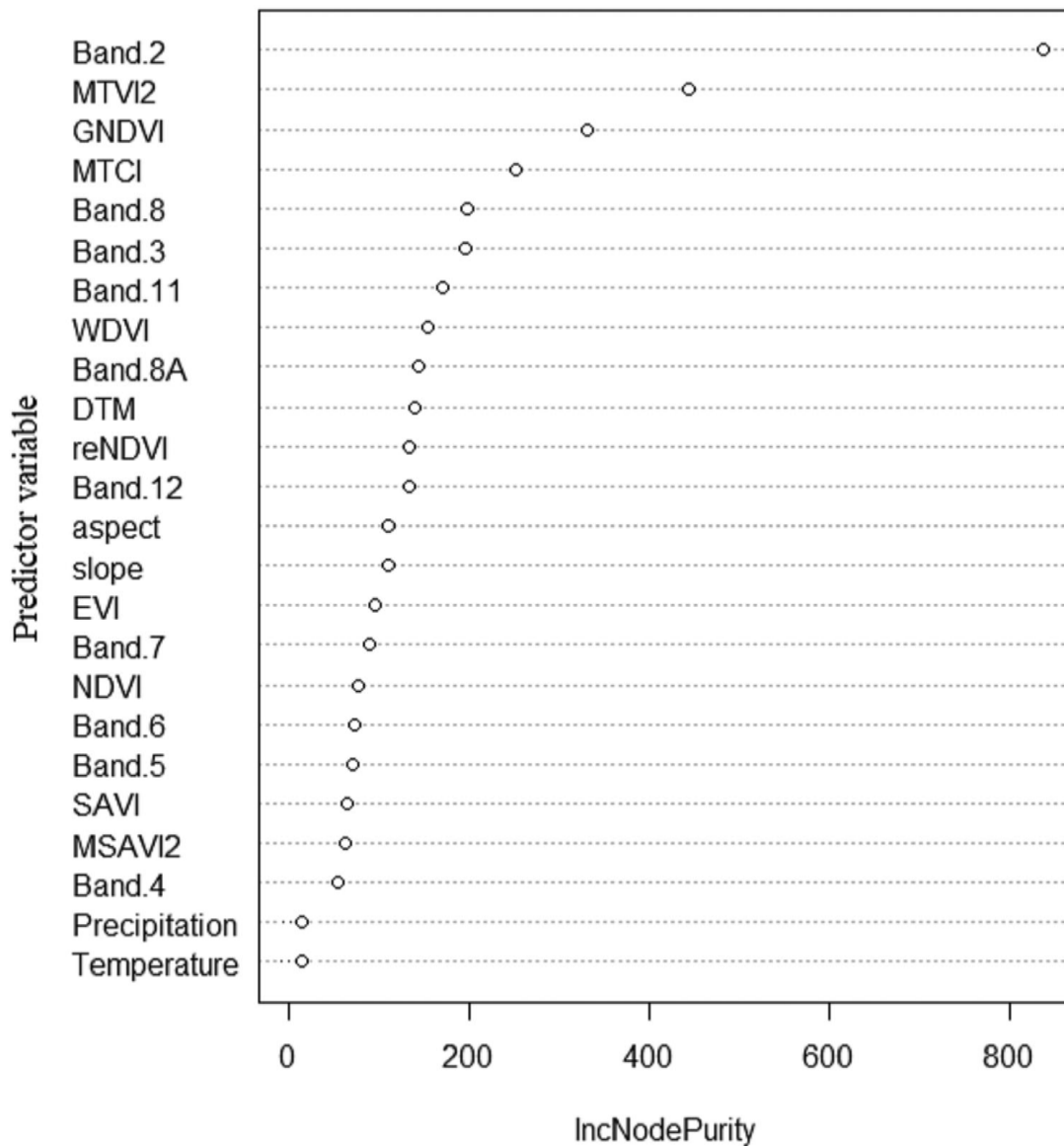
Figure 4. Variable importance list of all RFR model (S2 + VIs + Topographic + Climate).

more efficient than SVM and boosted regressions when dealing with large inputs and noisy datasets that may overfit and start modeling the noise (Segarra, Luis Araus, and Kefauver 2022; Adugna, Xu, and Fan 2022; Kuradusenge et al. 2023). Recent studies have proven the superiority of RFR to other algorithms for wheat yield estimation (Hunt et al. 2019). The results of this study showed that the RFR model could estimate the wheat yield at the field scale with RMSEs of 0.301 and 0.274 t/ha when using PS and S2 data, respectively (Figures 5 and 9). The developed model estimated the wheat yield with much higher accuracy and consistency than the model of (Segarra, Luis Araus, and Kefauver 2022). They also estimated the wheat yield at the field level by using an RFR model with VIs and LAI retrieved from S2 data but realized an RSME of 0.74 t/ha.

RFR can be used to effectively was used to determine relevant predictor variables. It showed that PS VIs including reNDVI, SAVI, EVI, WDVI and DEM were the top five important variables (Figure 3). Because reNDVI is critical to detect the amount of chlorophyl in the plants and is ideally applied in the mid-to-late growing season when wheat was the milk and dough stage (Saad El Imanni, Harti, and Iysaouy 2022). SAVI, EVI and WDVI are sensitive in vegetation structures and LAI, which were also influential in predicting crop yield (Kaplan et al. 2023). S2 Band 2 were the best predictor due to

Observed

Field 1   Field 2

Predicted

Yield
(tonnes/ha)

4   10
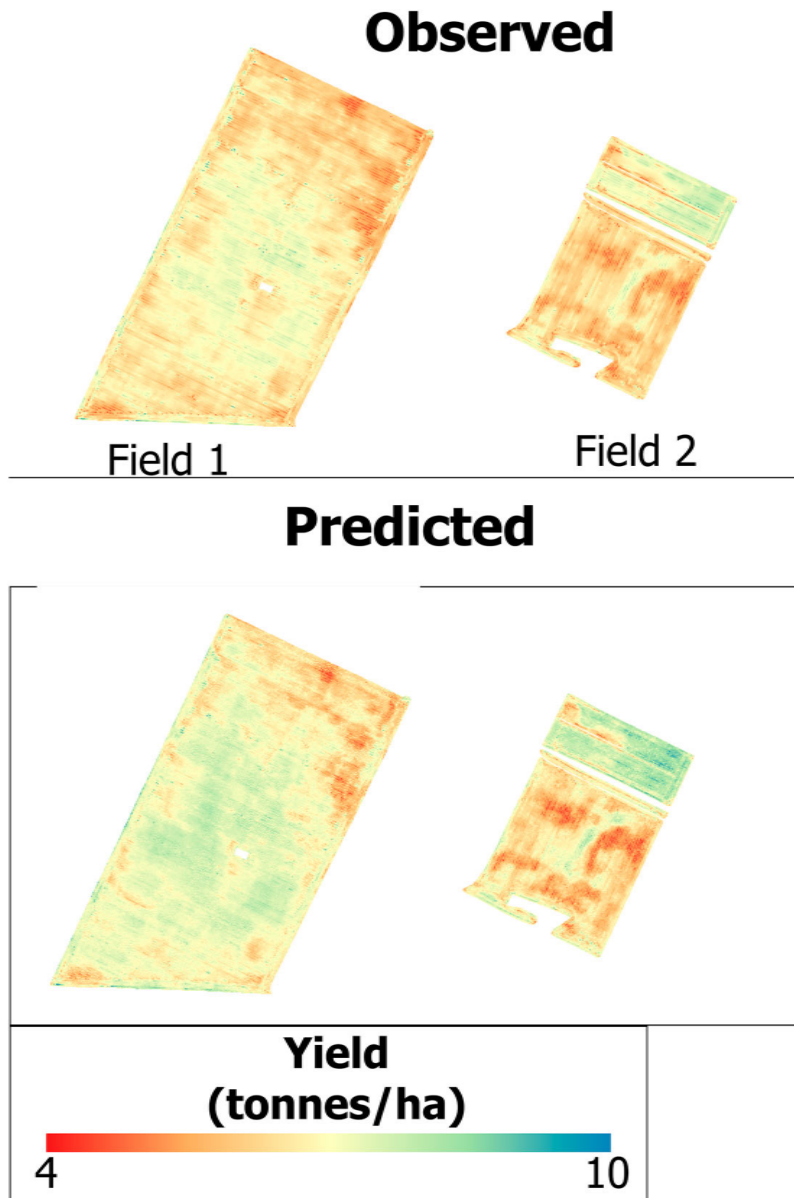
Figure 5. Pixel-level yield estimation with the RFR model using PS data.

chlorophyl pigments selectively absorbing blue (400–500 nm) light for photosynthesis which has a strong connection with yield (Figure 4). At the canopy scale, MTVI2 can detect leaf chlorophyll content but is relatively insensitive to LAI. GNDVI was the third most important variable in the S2-based model that is associated with the crop canopy water and nitrogen uptake which is successful for crop yield estimation. DTM was among the ten most important variables due to its high resolution which can capture different local roughness. The MTCI was also the most important feature due to the sensitivity of canopy chlorophyll content. The climatic data had a coarse pixel size of 4 km, resulting in almost regionally identical values across fields. Thus, it contributed the least to the prediction accuracy.

4.2. Phenological fitting and monitoring

We applied VIs and spectral data from the peak phenological stages to build the estimation model. To do this, we determined the importance of temporal variations in the sensed information to accurate yield estimation. The data indicated that peak wheat growth occurred between May and June,
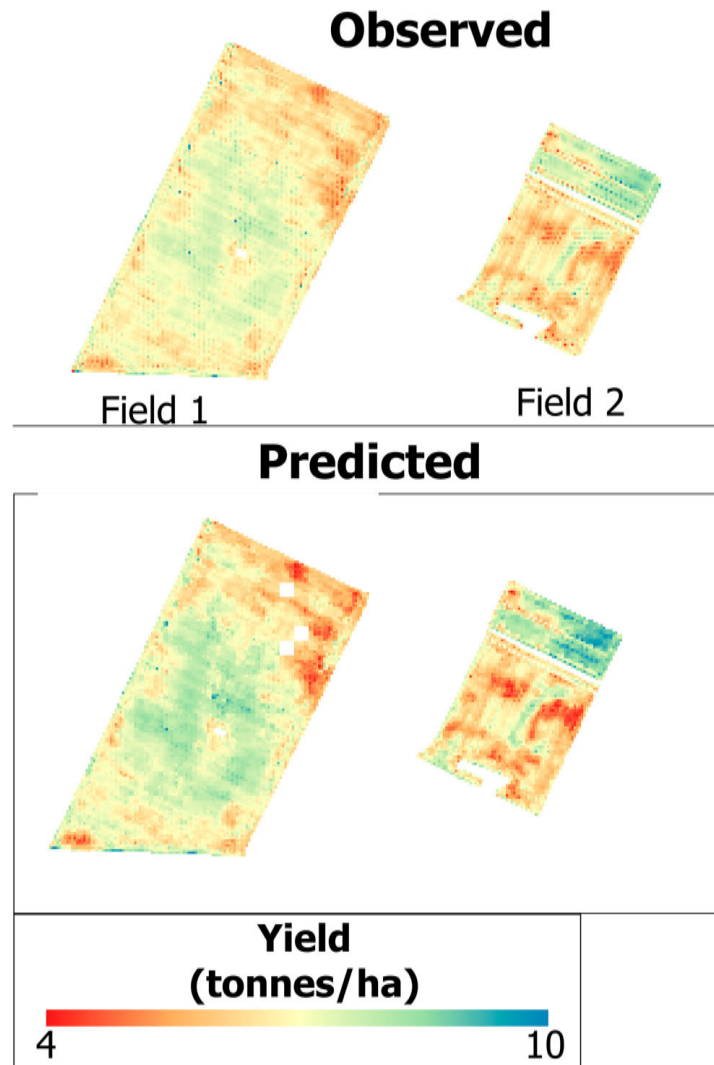
**Figure 6.** Pixel-level yield estimation with the RFR model using S2 data.

which corresponded to the botting, flowering, and fertilization stage and the milk and dough stage (BBCH-scale: 41–83). In the model training, RMSE values of 0.325–0.411 t/ha were obtained (Tables 9 and 10). Multiple studies have reported that peak VI values provided better yield estimation accuracy (Skakun et al. 2019; Ziliani et al. 2022). The PS and S2 satellite images acquired on June 17 and 20, respectively, helped the RFR model accurately estimate the yield. This result supports the results of (Zhao et al. 2020), who estimated the wheat yield at the field scale by combining VIs with S2 data and realized an RMSE of 0.64 t/ha.

4.3. PlanetScope vs. Sentinel-2

In this study, we examined the potential application of multispectral satellite data to estimating the wheat grain yield at the field scale. We compared the performances of the RFR model utilizing PS and S2 data considering the differences between the sensors and the tradeoff between accuracy and costs. The training results showed that the S2-based RFR model estimated the yield with slightly higher accuracy (RMSE = 0.325 t/ha, R^2 = 0.81) than the higher-resolution PS-based model (RMSE = 0.336 t/ha, R^2 = 0.79) using only basic spectral bands (Tables 7 and 8). The validation results also showed that the PS bands were not always better at explaining the wheat yield variability than the S2 data. Although the PS data had a higher temporal and
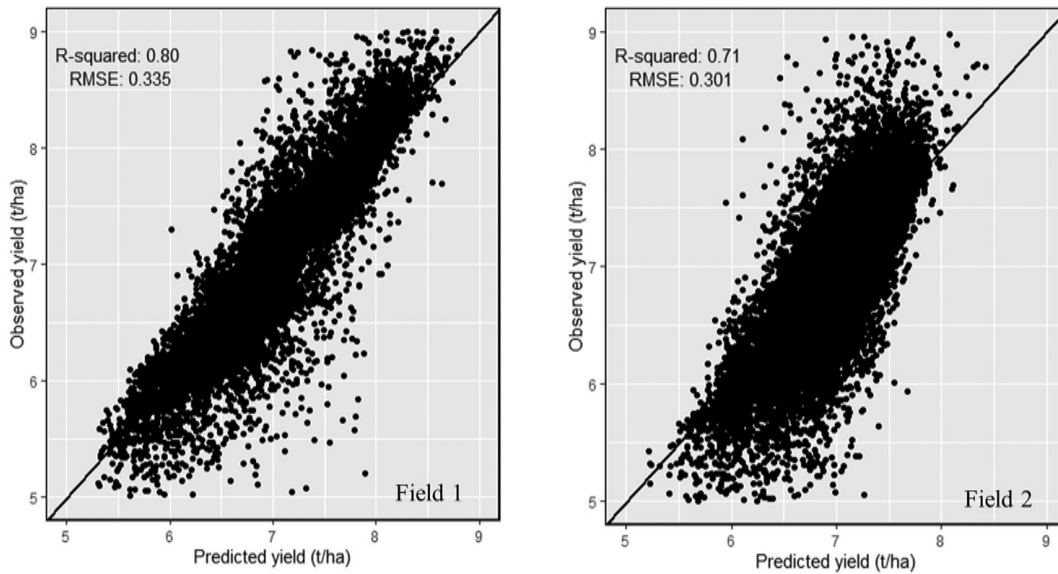
Figure 7. Comparison between the actual yields of the validation fields and the estimated yields using the RFR model with PS data, VIs, and environmental variables.
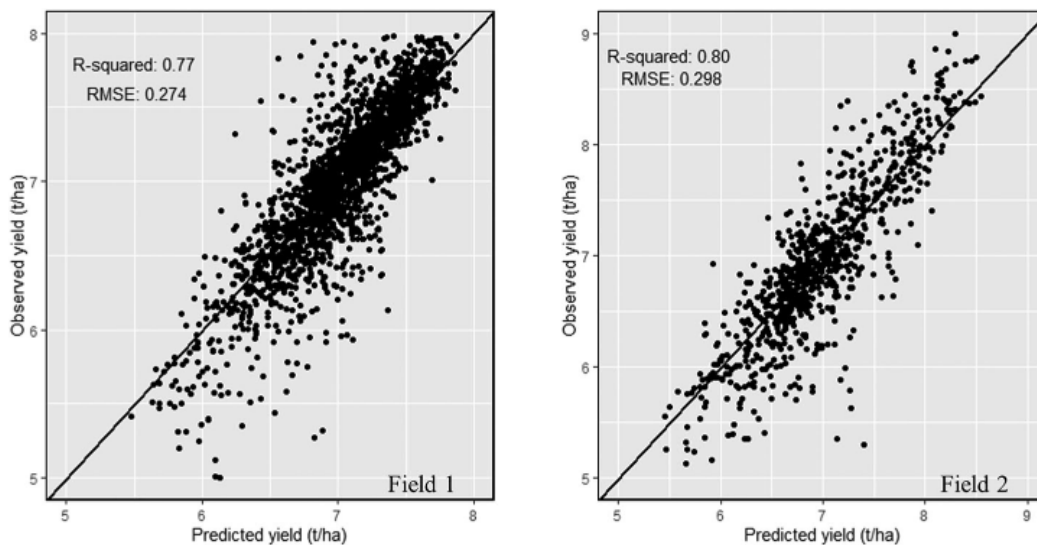

Figure 8. Comparison between the actual yields of the validation fields and the estimated yields using the RFR model with S2 data, VIs, and environmental variables.

spatial resolution, it had a lower radiometric resolution than the S2 data. The lack of two additional red edge and SWIR bands in the PS data may also be a factor. However, the PS products are provided almost daily, which provides an opportunity to improve the prediction accuracy and to promote digital agriculture for crop modeling and yield estimation (Ziliani et al. 2022). However, many studies have demonstrated that satellite imagery with high spatiotemporal resolution (e.g. S2 and Landsat 8) often fail to resolve the field-scale yield variability that is important for precision agriculture applications, especially for smaller fields (i.e. < 2 ha) (Jain et al. 2017). For example, Landsat 8 images can contain different spectral information because of the coarse 30-m spatial resolution.

VIs derived from the satellite image data were added as additional information for the model to analyze. Previous studies have developed empirical connections between the crop yield and VIs or
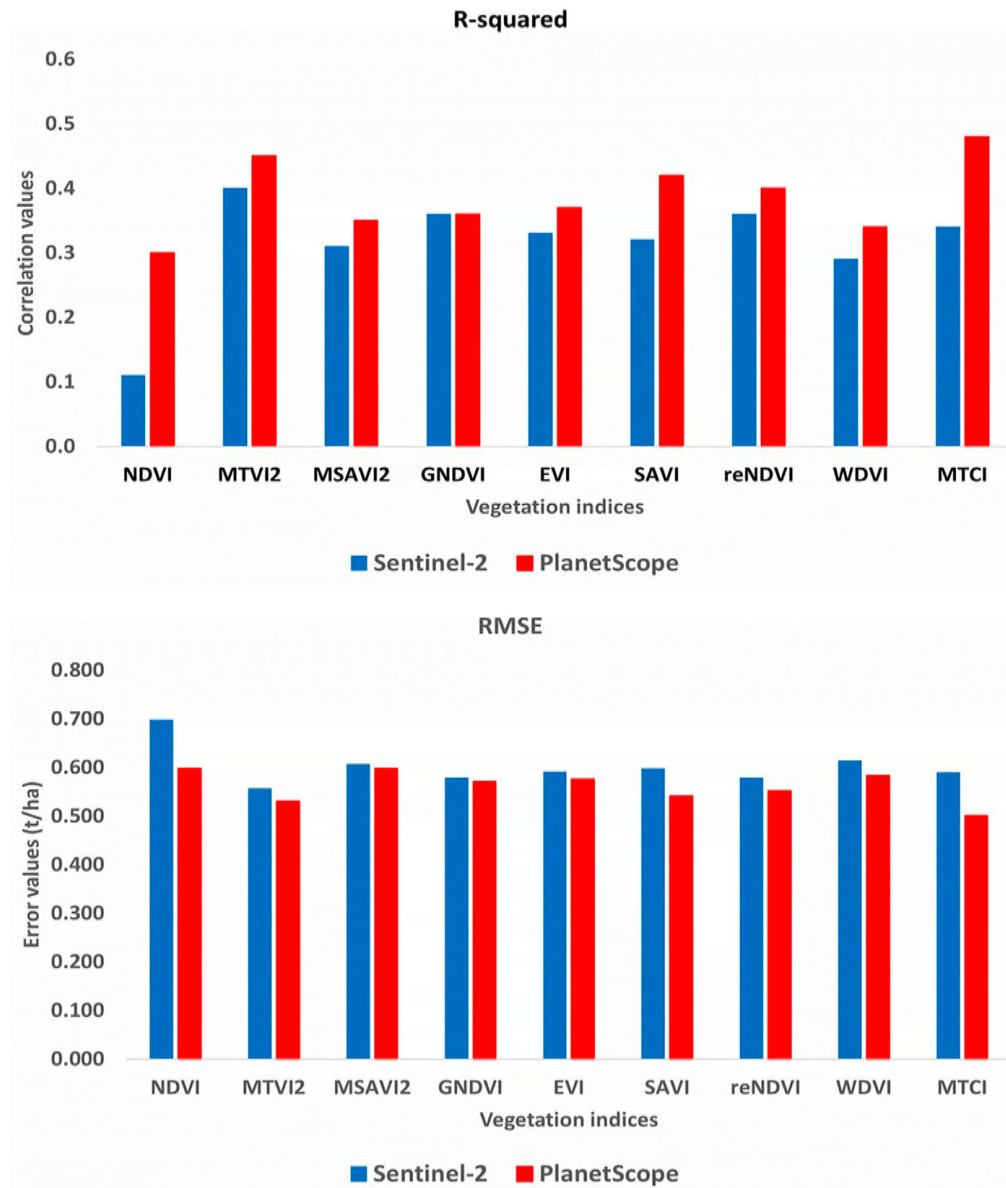
Figure 9. Evaluation of the VIs in prediction yield.

biophysical factors such as the LAI to estimate the yield in large plots with homogenous crops (Segarra, Luis Araus, and Kefauver 2022; Skakun et al. 2019). In our study, we found that using VIs and basic spectral bands together sometimes improved the estimation accuracy but not always. Some studies have found that calculating VIs separately did not improve the yield estimation accuracy (Hunt et al. 2019). Therefore, the RFR model can determine pertinent data for yield estimation from the satellite bands rather than relying on VIs.

We evaluated the impact of environmental data when combined with the basic spectral bands and VIs for regression analysis. The estimation accuracy was highest when the environmental data were combined with the PS and S2 data, and the developed model outperformed previously established models. Various previous studies have combined environmental data with satellite data to support crop yield estimation, (Burt 2012; Hunt et al. 2019; Schwalbert et al. 2020; Manivasagam et al. 2021). In the training results, the highest yield estimation accuracy was obtained by introducing environmental data to the PS data (RMSE = 0.287 t/ha, R^2 = 0.84) (Table 11). We supposed that the findings of this study and the achieved results can contribute to precision agriculture. First, the modeling

algorithm is important for accurate yield estimation. Second, the availability of the time-series satellite imagery connected to phenology, followed by the processing of spectral and combine harvester data. Lastly, the integration of multisource data can significantly increase yield estimation accuracy at the field level.

5. Conclusion

In this study, we evaluated the feasibility of applying PS and S2 data to wheat yield estimation at the field scale. The validation results showed that the RFR model achieved an RMSE of 0.413–0.438 t/ha with 3-m PS data and 0.301–0.338 t/ha with 10-m S2 data. Adding topographic and climatic datasets to the PS and S2 data further improved the yield estimation accuracy with RMSE values of 0.301–0.335 and 0.274–0.298 t/ha, respectively. To our knowledge, topographic and climatic variables have not previously been combined with high-resolution satellite images such as PS for wheat yield mapping. In addition, few studies have combined weather data with VIs derived from satellite data. This is also the first study to use eight bands of PS image data to estimate the wheat yield at the field level. Only a limited number of studies have assessed multiple sources of satellite data for wheat yield estimation at the field scale. Our results offer new developments in the methodology for field-scale wheat yield estimation. When comparing the time series of phenological stages, we found that the yield estimation accuracy was highest when data from the botting, flowering, and fertilization stage and the milk and dough stage (BBCH-scale: 41–83) were used, which corresponded to when the crops reached their maximum growth. The most suitable period for estimating the field-scale wheat yield was approximately 30–40 days before harvest during the milk and dough stage. The developed model can be applied to other crops and locations if suitable training data are available.

Data availability statement

All datasets R code used in this study are available from Nizom Farmonov (farmonov.nizom@stud.u-szeged.hu) upon reasonable request.

ORCID

Nizom Farmonov ⓘ http://orcid.org/0000-0002-2491-9340
Khilola Amankulova ⓘ http://orcid.org/0000-0001-6562-5616
József Szatmári ⓘ http://orcid.org/0000-0002-7896-3363
László Mucsi ⓘ http://orcid.org/0000-0002-5807-3742

References

Abatzoglou, John T., Solomon Z. Dobrowski, Sean A. Parks, and Katherine C. Hegewisch. 2018. "TerraClimate, a High-Resolution Global Dataset of Monthly Climate and Climatic Water Balance from 1958–2015." *Scientific Data* 5 (1): 170191. doi:10.1038/sdata.2017.191.

Adugna, Tesfaye, Wenbo Xu, and Jinlong Fan. 2022. "Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images." *Remote Sensing* 14 (3): 574. doi:10.3390/rs14030574.

Balogh, Szabolcs, and Tibor József Novák. 2020. "Trends and Hotspots in Landscape Transformation Based on Anthropogenic Impacts on Soil in Hungary, 1990–2018." *Hungarian Geographical Bulletin* 69 (4): 349–361. doi:10.15201/hungeobull.69.4.2.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.

Burt, P. J. A. 2012. "M. A. Sutton, C. M. Howard, J. W. Erisman, G. Billen, A. Bleeker, P. Grennfelt, H. Van Grinsven and B. Grizzetti (Eds), 2011. The European Nitrogen Assessment: Sources, Effects and Policy Perspectives, Cambridge University Press, UK. ISBN: 978-1-107-00612." *Meteorological Applications* 19 (1): E2–E2. doi:10.1002/met.1290.

Clevers, J. G. P. W. 1989. "Application of a Weighted Infrared-Red Vegetation Index for Estimating Leaf Area Index by Correcting for Soil Moisture." *Remote Sensing of Environment* 29 (1): 25–37. doi:10.1016/0034-4257(89)90076-X.

Dash, J., and P. J. Curran. 2007. "Evaluation of the MERIS Terrestrial Chlorophyll Index (MTCI)." *Advances in Space Research* 39 (1): 100–104. doi:10.1016/j.asr.2006.02.034.

Fawagreh, Khaled, Mohamed Medhat Gaber, and Eyad Elyan. 2014. "Random Forests: From Early Developments to Recent Advancements." *Systems Science & Control Engineering* 2 (1): 602–609. doi:10.1080/21642583.2014.956265.

Gitelson, Anatoly, and Mark N. Merzlyak. 1994. "Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus Hippocastanum L. and Acer Platanoides L. Leaves. Spectral Features and Relation to Chlorophyll Estimation." *Journal of Plant Physiology* 143 (3): 286–292. doi:10.1016/S0176-1617(11)81633-0.

Gitelson, A. A., Y. J. Kaufman, and M. N. Merzlyak. 1996. "Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS." *Remote Sensing of Environment* 58: 289–298. doi:10.1016/S0034-4257(96)00072-7

Haboudane, D. 2004. "Hyperspectral Vegetation Indices and Novel Algorithms for Predicting Green LAI of Crop Canopies: Modeling and Validation in the Context of Precision Agriculture." *Remote Sensing of Environment* 90 (3): 337–352. doi:10.1016/j.rse.2003.12.013.

Han, Jichong, Zhao Zhang, Juan Cao, Yuchuan Luo, Liangliang Zhang, Ziyue Li, and Jing Zhang. 2020. "Prediction of Winter Wheat Yield Based on Multi-Source Data and Machine Learning in China." *Remote Sensing* 12 (2): 236. doi:10.3390/rs12020236.

Huete, A. R. 1988. "A Soil-Adjusted Vegetation Index (SAVI)." *Remote Sensing of Environment* 25 (3): 295–309. doi:10.1016/0034-4257(88)90106-X.

Huete, A., K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira. 2002. "Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices." *Remote Sensing of Environment* 83 (1–2): 195–213. doi:10.1016/S0034-4257(02)00096-2.

Hunt, Merryn L., George Alan Blackburn, Luis Carrasco, John W. Redhead, and Clare S. Rowland. 2019. "High Resolution Wheat Yield Mapping Using Sentinel-2." *Remote Sensing of Environment* 233 (November): 111410. doi:10.1016/j.rse.2019.111410.

Igrejas, Gilberto, and Gérard Branlard. 2020. "The Importance of Wheat." In *Wheat Quality For Improving Processing And Human Health*, edited by Gilberto Igrejas, Tatsuya M. Ikeda, and Carlos Guzmán, 1–7. Cham: Springer International Publishing. doi:10.1007/978-3-030-34163-3_1.

Ittersum, Martin K. van. 2016. "Crop Yields and Global Food Security. Will Yield Increase Continue to Feed the World?" *European Review of Agricultural Economics* 43 (1): 191–192. doi:10.1093/erae/jbv034.

Jackson, T. 2004. "Vegetation Water Content Mapping Using Landsat Data Derived Normalized Difference Water Index for Corn and Soybeans." *Remote Sensing of Environment* 92: 475–482. doi:10.1016/j.rse.2003.10.021.

Jain, M., Balwinder Singh, A. A. K. Srivastava, R. K. Malik, A. J. McDonald, and D. B. Lobell. 2017. "Using Satellite Data to Identify the Causes of and Potential Solutions for Yield Gaps in India's Wheat Belt." *Environmental Research Letters* 12 (9): 0094011. doi:10.1088/1748-9326/aa8228.

Jeong, Jig Han, Jonathan P. Resop, Nathaniel D. Mueller, David H. Fleisher, Kyungdahm Yun, Ethan E. Butler, Dennis J. Timlin, et al. 2016. "Random Forests for Global and Regional Crop Yield Predictions." Edited by Jose Luis Gonzalez-Andujar. *PLOS ONE* 11 (6): e0156571. doi:10.1371/journal.pone.0156571.

Kaplan, Gregoriy, Lior Fine, Victor Lukyanov, Nitzan Malachy, Josef Tanny, and Offer Rozenstein. 2023. "Using Sentinel-1 and Sentinel-2 Imagery for Estimating Cotton Crop Coefficient, Height, and Leaf Area Index." *Agricultural Water Management* 276 (February): 108056. doi:10.1016/j.agwat.2022.108056.

Kharel, Tulsi P., Angel Maresma, Karl J. Czymmek, Erasmus K. Oware, and Quirine M. Ketterings. 2019. "Combining Spatial and Temporal Corn Silage Yield Variability for Management Zone Development." *Agronomy Journal* 111 (6): 2703–2711. doi:10.2134/agronj2019.02.0079.

Kumhálová, Jitka, and Štěpánka Matějková. 2017. "Yield Variability Prediction by Remote Sensing Sensors with Different Spatial Resolution." *International Agrophysics* 31 (2): 195–202. doi:10.1515/intag-2016-0046.

Kuradusenge, Martin, Eric Hitimana, Damien Hanyurwimfura, Placide Rukundo, Kambombo Mtonga, Angelique Mukasine, Claudette Uwitonze, Jackson Ngabonziza, and Angelique Uwamahoro. 2023. "Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize." *Agriculture* 13 (1): 225. doi:10.3390/agriculture13010225.

Lancashire, Peter D., H. Bleiholder, T. Van Den Boom, P. Langelüddeke, R. Stauss, Elfriede Weber, and A. Witzenberger. 1991. "A Uniform Decimal Code for Growth Stages of Crops and Weeds." *Annals of Applied Biology* 119 (3): 561–601. doi:10.1111/j.1744-7348.1991.tb04895.x.

Liaw, A., and M. Wiener. 2002. "Classification and Regression by Random Forest." *R News* 2: 18-22.

Manivasagam, V. S., Yuval Sadeh, Gregoriy Kaplan, David J. Bonfil, and Offer Rozenstein. 2021. "Studying the Feasibility of Assimilating Sentinel-2 and PlanetScope Imagery Into the SAFY Crop Model to Predict Within-Field Wheat Yield." *Remote Sensing* 13 (12): 2395. doi:10.3390/rs13122395.

Matcham, Emma G., Filipe Matias, Brian D. Luck, and Shawn P. Conley. 2022. "Filtering, Editing, and Cropping Yield Maps in a R Environment with the Package CleanRfield." *Agronomy Journal* 114 (3): 1672–1679. doi:10.1002/agj2.21055.

Maya Gopal, P. S., and R. Bhargavi. 2019. "Performance Evaluation of Best Feature Subsets for Crop Yield Prediction Using Machine Learning Algorithms." *Applied Artificial Intelligence* 33 (7): 621–42. doi:10.1080/08839514.2019.1592343.

Planet Team. 2017. *Planet Application Program Interface: In Space for Life on Earth.* San Francisco, CA. https://api.planet.com.

Rehfeldt, Gerald E., Nicholas L. Crookston, Cuauhtémoc Sáenz-Romero, and Elizabeth M. Campbell. 2012. "North American Vegetation Model for Land-Use Planning in a Changing Climate: A Solution to Large Classification Problems." *Ecological Applications* 22 (1): 119–41. doi:10.1890/11-0495.1.

Rosegrant, Mark W., and Sarah A. Cline. 2003. "Global Food Security: Challenges and Policies." *Science* 302 (5652): 1917–1919. doi:10.1126/science.1092958.

Rouse, J. W., R. H. Haas, J. A. Schell, and D. W. Deering. 1973. "Monitoring Vegetation Systems in the Great Plains with ERTS (Earth Resources Technology Satellite)." *Proceedings of 3rd Earth Resources Technology Satellite Symposium*, Greenbelt, 10-14 December, SP-351, 309-317.

Ruml, Mirjana, and Todor Vulic. 2005. "Importance of Phenological Observations and Predictions in Agriculture." *Journal of Agricultural Sciences, Belgrade* 50 (2): 217–225. doi:10.2298/JAS0502217R.

Saad El Imanni, Hajar, Abderrazak El Harti, and Lahcen El Iysaouy. 2022. "Wheat Yield Estimation Using Remote Sensing Indices Derived from Sentinel-2 Time Series and Google Earth Engine in a Highly Fragmented and Heterogeneous Agricultural Region." *Agronomy* 12 (11): 2853. doi:10.3390/agronomy12112853.

Schwalbert, Raí A., Telmo Amado, Geomar Corassa, Luan Pierre Pott, P. V. Vara Prasad, and Ignacio A. Ciampitti. 2020. "Satellite-Based Soybean Yield Forecast: Integrating Machine Learning and Weather Data for Improving Crop Yield Prediction in Southern Brazil." *Agricultural and Forest Meteorology* 284 (April): 107886. doi:10.1016/j.agrformet.2019.107886.

Segarra, Joel, Jose Luis Araus, and Shawn C. Kefauver. 2022. "Farming and Earth Observation: Sentinel-2 Data to Estimate Within-Field Wheat Grain Yield." *International Journal of Applied Earth Observation and Geoinformation* 107 (March): 102697. doi:10.1016/j.jag.2022.102697.

Shanahan, J. F., J. S. Schepers, D. D. Francis, G. E. Varvel, W. W. Wilhelm, J. M. Tringe, M. R. Schlemmer, and D. J. Major. 2001. "Use of Remote-Sensing Imagery to Estimate Corn Grain Yield." *Agronomy Journal* 93: 583–589. doi:10.2134/agronj2001.933583x.

Skakun, Sergii, Eric Vermote, Belen Franch, Jean-Claude Roger, Nataliia Kussul, Junchang Ju, and Jeffrey Masek. 2019. "Winter Wheat Yield Assessment from Landsat 8 and Sentinel-2 Data: Incorporating Surface Reflectance, Through Phenological Fitting, Into Regression Yield Models." *Remote Sensing* 11 (15): 1768. doi:10.3390/rs11151768.

Smith, Paul F., Siva Ganesh, and Ping Liu. 2013. "A Comparison of Random Forest Regression and Multiple Linear Regression for Prediction in Neuroscience." *Journal of Neuroscience Methods* 220 (1): 85–91. doi:10.1016/j.jneumeth.2013.08.024.

Singh, Balraj, Parveen Sihag, and Karan Singh. 2017. "Modelling of Impact of Water Quality on Infiltration Rate of Soil by Random Forest Regression." *Modeling Earth Systems and Environment* 3 (3): 999–1004. doi:10.1007/s40808-017-0347-3.

Tucker, Compton J. 1979. "Red and Photographic Infrared Linear Combinations for Monitoring Vegetation." *Remote Sensing of Environment* 8 (2): 127–150. doi:10.1016/0034-4257(79)90013-0.

Vijayasekaran, D. 2019. "SEN2-AGRI – CROP TYPE MAPPING PILOT STUDY USING SENTINEL-2 SATELLITE IMAGERY IN INDIA." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-3/W6 (July): 175–180. doi:10.5194/isprs-archives-XLII-3-W6-175-2019.

Qi, J., A. Chehbouni, A. R. Huete, Y. H. Kerr, and S. Sorooshian. 1994. "A Modified Soil Adjusted Vegetation Index." *Remote Sensing of Environment* 48: 119–126. doi:10.1016/0034-4257(94)90134-1

Xie, Xuefeng, Tao Wu, Ming Zhu, Guojun Jiang, Yan Xu, Xiaohan Wang, and Lijie Pu. 2021. "Comparison of Random Forest and Multiple Linear Regression Models for Estimation of Soil Extracellular Enzyme Activities in Agricultural Reclaimed Coastal Saline Land." *Ecological Indicators* 120 (January): 106925. doi:10.1016/j.ecolind.2020.106925.

Zhao, Yan, B Potgieter Andries, Miao Zhang, Bingfang Wu, and Graeme L Hammer. 2020. "Predicting Wheat Yield at the Field Scale by Combining High-Resolution Sentinel-2 Satellite Imagery and Crop Modelling." *Remote Sensing* 12 (6): 1024. doi:10.3390/rs12061024.

Ziliani, Matteo G., Muhammad U. Altaf, Bruno Aragon, Rasmus Houborg, Trenton E. Franz, Yang Lu, Justin Sheffield, Ibrahim Hoteit, and Matthew F. McCabe. 2022. "Early Season Prediction of Within-Field Crop Yield Variability by Assimilating CubeSat Data Into a Crop Model." *Agricultural and Forest Meteorology* 313 (February): 108736. doi:10.1016/j.agrformet.2021.108736.