Precision Agriculture (2022) 23:1093–1123
https://doi.org/10.1007/s11119-022-09876-5

Strategies for monitoring within-field soybean yield using Sentinel-2 Vis-NIR-SWIR spectral bands and machine learning regression methods

L. G.T. Crusiol[1,2] · Liang Sun[1] · R. N.R. Sibaldelli[3] · V. Felipe Junior[4] · W. X. Furlaneti[4] · R. Chen[1] · Z. Sun[1] · D. Wuyun[1] · Z. Chen[5] · M. R. Nanni[2] · R. H. Furlanetto[2] · E. Cezar[2] · A. L. Nepomuceno[3] · J. R.B. Farias[3]

Accepted: 14 January 2022 / Published online: 19 April 2022

Abstract
Soybean crop plays an important role in world food production and food security, and agricultural production should be increased accordingly to meet the global food demand. Satellite remote sensing data is considered a promising proxy for monitoring and predicting yield. This research aimed to evaluate strategies for monitoring within-field soybean yield using Sentinel-2 visible, near-infrared and shortwave infrared (Vis/NIR/SWIR) spectral bands and partial least squares regression (PLSR) and support vector regression (SVR) methods. Soybean yield maps (over 500 ha) were recorded by a combine harvester with a yield monitor in 15 fields (3 farms) in Paraná State, southern Brazil. Sentinel-2 images (spectral bands and 8 vegetation indices) across a cropping season were correlated to soybean yield. Information pooled across the cropping season presented better results compared to single images, with best performance of Vis/NIR/SWIR spectral bands under PLSR and SVR. At the grain filling stage, field-, farm- and global-based models were evaluated and presented similar trends compared to leaf-based hyperspectral reflectance collected at the Brazilian National Soybean Research Center. SVR outperformed PLSR, with a strong correlation between observed and predicted yield. For within-field soybean yield mapping, field-based SVR models (developed individually for each field) presented the highest accuracies. The results obtained demonstrate the possibility of developing within-field yield prediction models using Sentinel-2 Vis/NIR/SWIR bands through machine learning methods.

Keywords Yield prediction · Yield mapping · Partial least squares regression · Support vector regression · Multispectral image · Multitemporal data

Extended author information available on the last page of the article

Introduction

Considering the projections of world population, expected to be over nine billion people by 2050 (FAO, 2018), agricultural production should be increased accordingly to meet the global food demand. Therefore, to guarantee high levels of productivity while preserving the environment, agricultural land use and agronomic management practices must follow sustainable practices, using precise and time-efficient information about crop spatial distribution and development conditions. In this scenario, remote sensing has the potential to provide accurate spatial information about agricultural systems, contributing to better site-specific management of agronomic practices, with effects over financial market and strategic planning of governmental and corporative policies, such as supply regulation and food security (Gusso & Ducati 2012; Silva Junior et al., 2017).

Soybean plays an important role in world food production and food security, and Brazil is responsible for more than one third (over 135 Mt—CONAB, 2021) of soybean produced worldwide (362 Mt—USDA, 2021). To address site-specific and time-efficient crop management, remote sensing data has been considered a promising proxy for monitoring and predicting soybean yield at multiple levels of data acquisition: satellite-based (Bolton & Friedl 2013; Dado et al., 2020), UAV-based (Zhang et al., 2019; Silva et al., 2020) and field-based (Christenson et al., 2016; Crusiol et al., 2017a, 2021b; Carneiro et al., 2020) and at multiple scales, varying from limited spatial extents to county- or state-level (Dado et al., 2020). However, the monitoring of within-field soybean yield still faces limitations and there is a substantial need for more research due to the low availability of high spatial resolution yield data (acquired by combine harvester with yield monitors, despite their increasing use in agricultural activities) and due to the complex relationships among crop growth processes and their biotic and abiotic stresses within each field (Dado et al., 2020, Kross et al., 2020).

Usually, most yield prediction models using satellite remote sensed data are developed using vegetation indices (VIs—considered as the key indicator of yield), based on their direct relationship with biomass and the indirect relationship between biomass and yield, which may not always represent the variability in yield caused by stresses across a cropping season (Sakamoto 2020). For this reason, the use of reflectance from multiple spectral bands, analyzed by machine learning algorithms, has demonstrated higher accuracy than VIs for yield prediction in soybean crop systems using multispectral satellite data (Dado et al., 2020) and aerial multispectral imagery (Leon et al., 2003).

Machine learning algorithms are useful to identify patterns in datasets, capturing complex relationships among driving variables and delivering higher accuracy compared to conventional correlations, being considered an important tool in decision support systems for agricultural management (Kayad et al., 2019; Zhai et al., 2020). Based on remote sensing data, the advantage of machine learning methods for regression tasks, either linear (e.g., partial least squares regression) or non-linear (e.g., support vector regression), relies on the possibility of using two or more spectral variables (e.g., spectral bands or vegetation indices) to predict a key parameter of crop development (e.g., yield). Unlike the traditional univariate regression methods, in which one spectral variable (e.g., spectral band or vegetation index) is used to predict yield, machine learning regression methods enable the development of prediction models using several spectral bands or vegetation indices acquired from the target area, or even their combination (spectral bands and vegetation indices), contributing to

the better characterization of the crop development condition across different wavelengths. Considering the important role of time for crop monitoring, machine learning regression methods enable the use, in the same model, of spectral information, such as image spectral bands and derived vegetation indices. This spectral information can be acquired on different days across the cropping season, which has the potential of better characterizing the timing of occurrence of biotic and abiotic factors that can influence yield values. In this context, since machine learning regression methods gather information from different spectral bands and vegetation indices at different times across a cropping season, there is a large potential for the developed models to explain crop variability and its yield.

According to Gao et al. (2018), yield prediction through remote sensed data is usually based on a single day observation, which poses limitations on the understanding of physiological trends in crop systems and their relationship to yield. Hence, the R5 phenological stage of soybean crop (grain filling stage, Fehr & Caviness, 1977) has been described as the most suitable for yield prediction using remote sensed data (Sakamoto 2020; Crusiol et al., 2021a). However, combining information from different times across a cropping season can contribute to overcoming such limitations and has demonstrated high accuracy in yield modelling (Hunt et al., 2019; Dado et al., 2020; Gómez et al., 2019).

Under machine learning perspectives, Sentinel-2 imagery (European Space Agency—ESA) has been increasingly used for yield monitoring due to its high spatial, temporal and spectral resolutions. With a revisit time of five days, 13 spectral bands (from visible to short-wave infrared wavelengths) and spatial resolution between 10 and 60 m (depending on the spectral band), Sentinel-2 has made significant spectral data available, providing a better understanding of crop dynamics and contributing to the development of crop yield prediction models (Di Gennaro et al., 2019; Segarra et al., 2020).

Although research into yield monitoring in several crop types using remote sensed data and machine learning algorithms has been carried out, the quantitative assessment of different input variables (spectral bands and vegetation indices from single or pooled images across a cropping season) for soybean within-field yield monitoring has been little discussed. Therefore, there is a great potential to further explore the combinations of input variables, contributing to the adoption of feasible strategies for yield prediction through satellite multispectral images. The present manuscript aimed to evaluate strategies for monitoring within field soybean yield using Sentinel-2 visible, near-infrared and shortwave infrared spectral bands and machine learning regressions methods. The specific goals addressed: (1) comparison between single vegetation indices and multiple spectral bands under machine learning algorithms for yield prediction; (2) the contribution of using all available Sentinel-2 images across a cropping season for yield prediction; (3) the development of a yield prediction model based on Sentinel-2 Vis-NIR-SWIR spectral bands at the R5 phenological stage; and (4) the mapping of within-field soybean yield based on Sentinel-2 images.

Materials and methods

Study areas

The study areas are located in Astorga and Mauá da Serra municipalities, situated in the north area of Paraná State, southern Brazil. Paraná state accounts for about 16% (over 20

Mt) of the national soybean production and for about 5% of all soybeans produced worldwide (SEAB, 2021, CONAB 2021, USDA 2021). Astorga and Mauá da Serra municipalities play an important role in soybean production in Paraná State.

Three soybean farms (Fig. 1), described in Table 1, were monitored in the 2019/2020 cropping season. Água-Viva and Tupinambá farms, named from now on as AV and TP farms respectively, are located in Astorga Municipality, while Mauá da Serra farm, named from now on as MS farm, is located in Mauá da Serra Municipality.

The climate of the experimental area is classified as Cfa according to Köppen climate classification, i.e., subtropical climate, with a mean temperature in the hottest month higher than 22 °C, and rainfall concentrated in the summer months, corresponding, therefore, to the period of soybean production, albeit with no defined dry season (Wrege et al., 2011; Alvares et al., 2013).

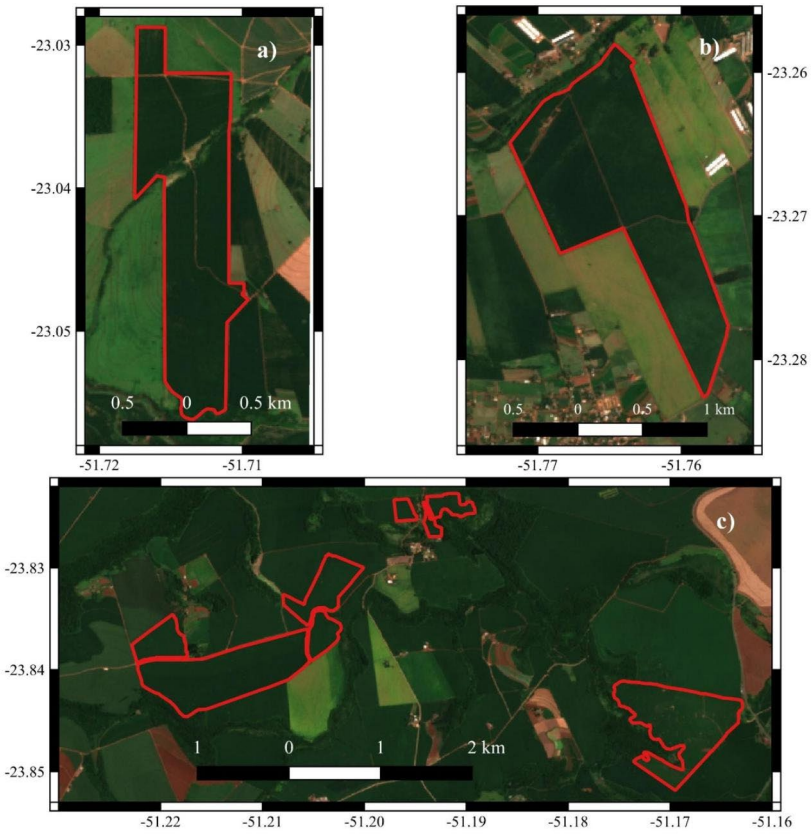


Fig. 1 Location of Água-Viva—AV (a), Tupinambá—TP (b) and Mauá da Serra—MS (c) farms using True Color Image (TCI), an RGB product from Sentinel-2 composed by bands 2 (blue), 3 (green) and 4 (red)

Table 1 Description of the monitored farms, fields, area, mean altitude, sowing and harvesting dates and cultivar

Farm	Field	Area (ha)	Mean altitude (m)	Sowing date	Harvesting date	Cultivar
AV	AV 1	46	443	From Oct. 30 to Nov. 02	From Mar. 04 to 22	BMX Fibra
	AV 2	55				BMX Garra
	AV 3	41				TMG 7067 IPRO
TP	TP 1	8	565	From Oct. 23 to 25	From Feb. 20 to Mar. 03	M6410 IPRO
	TP 2	43				TMG 7067 IPRO
	TP 3	48				M6410 IPRO
	TP 4	53				M6410 IPRO
MS	MS 1	12	898	From Oct. 22 to 27	From Mar. 11 to 16	BRS 511
	MS 2	15				BRS 511
	MS 3	16				BRS 511
	MS 4	24				BRS 511
	MS 5	48		From Nov. 06 to 08	From Mar. 17 to 23	M6410 IPRO
	MS 6	32		From Nov. 19 to 20	From Mar. 23 to 25	M6410 IPRO
	MS 7	9		From Dec. 01 to 02	From Mar. 27 to Apr. 03	BMX Compacta
	MS 8	65				BMX Compacta

Soybean yield data

Soybean yield data were obtained by cooperation with *Integrada Cooperativa Agroindustrial*, an agro-industrial cooperative of farmers that provides them technical assistance for crop production and logistic support for agricultural input supply and production trade. *Integrada Cooperativa Agroindustrial* is responsible for over 1% of all soybeans produced in Brazil and the monitored farms are included in the Precision Agriculture Program of the cooperative. Yield data were recorded by a combine harvester with yield monitor and the output data consisted of a point dataset containing multiple information about crop harvest, i.e., the geographic location of each point (acquired by the global navigation satellite system in the machinery, and recorded every 1 s), the speed of the harvester and the ground distance represented by each recorded point, header width, grain weight and grain moisture.

To eliminate the effects derived from null yield, overlapped points, harvested width less than the harvester header and speed variation, effects recognized to have a large impact in the delivered yield map (Vega et al., 2019), soybean yield was calculated and corrected to 13% grain moisture, as per Eq. 1.

$$GY = \frac{(100 - HGM)}{(100 - DGM)} \times HGW \times \frac{10{,}000}{(HW \times GD)} \tag{1}$$

In which GY is the grain yield (kg ha^{-1}), HGM the harvested grain moisture (%), DGM the desired grain moisture (%), HGW the harvested grain weight (kg), HW is the header width (m) and GD the ground distance represented by the point (m).

The yield dataset obtained for each field underwent statistical quality control using control charts with upper and lower control limits defined at –3σ and +3σ in relation to the

mean yield, where σ represents the standard deviation. Yield values above or below the control limits were considered as outlier and removed from the yield dataset, as suggested by Kross et al. (2020) and Carneiro et al. (2020). Using the software QGIS (QGIS, 2022), yield points were then rasterized to a 20 m cell grid, to meet the Sentinel-2 spatial resolution, as described in section 'Sentinel-2 data', and smoothed using a Gaussian filter with smoothing degree equal to 2.5σ.

Sentinel-2 data

Comprising the period of crop development, from November 2019 to March 2020, Sentinel-2 images were acquired from the Copernicus Open Hub at the Level 2 A Bottom-Of-Atmosphere (BOA) reflectance product. Sentinel-2 A and -2B, launched on 23 June 2015 and 7 March 2017 respectively (Gao et al., 2018), carry a Multi-Spectral Instrument (MSI) and provide a revisiting interval of 5 days. The spatial and spectral characteristics of MSI from Sentinel-2 A and -2B (ESA, 2021) are presented in supplementary table.

Nine spectral bands were used: bands 2, 3, 4, 5, 6, 7, 8a, 11 and 12. Band 8 was not used due to its overlapping with bands 7 and 8a, and bands 1, 9 and 10 were not used due to their coarse resolution (60 m) for within-field monitoring. Kayad et al. (2019), monitoring within-field maize yield, assessed the correlation between yield and Sentinel-2 bands 8 and 8a and observed similar and competitive results. Considering that most of the spectral bands used are originally acquired at 20 m resolution, bands 2, 3 and 4 were also analyzed at 20 m resolution (originally provided by Sentinel-2 Level 2 A product) to match the same spatial resolution of most spectral channels, as suggested by Dado et al. (2020). After cloud and shadow mask, only images, from each farm, containing the majority of their pixels free of cloud and shadow were used. Sentinel-2 images were processed using the software QGIS (QGIS, 2022).

Table 2 presents the images used in each field and the correspondent overpass date. The different number of available images for each field is related to the differences in sowing calendar (Table 1) and cloud cover on the overpass date. Despite the high revisiting frequency of Sentinel-2, the high cloud cover observed in the study area limited the number of available images from 4 to 7 images on each field. According to Eberhardt et al. (2016), Paraná State is negatively affected (from 40% up to 70%) by cloud cover on optical images, and Astorga and Mauá da Serra municipalities have an average (between the years of 2000 and 2014) of about only 30% of cloud-free images across the soybean cropping season. The negative effects of cloud cover for soybean monitoring in Paraná State have also been reported by Sugawara et al. (2008) and Crusiol et al. (2017b).

From the acquired Sentinel-2 images, vegetation indices (VIs) were calculated, as described in Table 3: blue normalized difference vegetation index (BNDVI), green normalized difference vegetation index (GNDVI), normalized difference vegetation index (NDVI), normalized difference red-edge index (NDRE), normalized difference infrared index (NDII), normalized difference infrared index 2 (NDII 2), enhanced vegetation index (EVI) and enhanced vegetation index 2 (EVI 2).

Table 2 Sentinel-2 images used for yield monitoring on each field

		Overpass date								
		Nov.	Dec.	Jan.		Feb.		Mar.		
Farm	Field	29	29	18	28	12	27	03	08	13
AV	AV 1	×	×	×	×	×				
	AV 2	×	×	×	×	×				
	AV 3	×	×	×	×	×				
TP	TP 1		×	×	×	×				
	TP 2		×	×	×	×				
	TP 3		×	×	×	×				
	TP 4		×	×	×	×				
MS	MS 1	×	×	×	×	×				
	MS 2	×	×	×	×	×				
	MS 3	×	×	×	×	×				
	MS 4	×	×	×	×	×				
	MS 5		×	×	×	×	×	×		
	MS 6			×	×	×	×	×	×	×
	MS 7			×	×	×	×	×	×	×
	MS 8			×	×	×	×	×	×	×

Table 3 Vegetation indices calculated from Sentinel-2 images

Index	Formula	Reference
BNDVI	$\mathrm{BNDVI} = \frac{(\rho\mathrm{NIR}-\rho\mathrm{Blue})}{(\rho\mathrm{NIR}+\rho\mathrm{Blue})}$	Wang et al. (2007)
GNDVI	$\mathrm{GNDVI} = \frac{(\rho\mathrm{NIR}-\rho\mathrm{Green})}{(\rho\mathrm{NIR}+\rho\mathrm{Green})}$	Gitelson et al. (1996)
NDVI	$\mathrm{NDVI} = \frac{(\rho\mathrm{NIR}-\rho\mathrm{Red})}{(\rho\mathrm{NIR}+\rho\mathrm{Red})}$	Rouse et al. (1974)
NDRE	$\mathrm{NDRE} = \frac{(\rho\mathrm{NIR}-\rho\mathrm{RedEdge})}{(\rho\mathrm{NIR}+\rho\mathrm{RedEdge})}$	Gitelson and Merzlyak (1994)
NDII	$\mathrm{NDII} = \frac{(\rho\mathrm{NIR}-\rho\mathrm{SWIR1600}\lambda)}{(\rho\mathrm{NIR}+\rho\mathrm{SWIR1600}\lambda)}$	Hardisky et al. (1983)
NDII 2	$\mathrm{NDII2} = \frac{(\rho\mathrm{NIR}-\rho\mathrm{SWIR2200}\lambda)}{(\rho\mathrm{NIR}+\rho\mathrm{SWIR2200}\lambda)}$	Hardisky et al. (1983)
EVI	$\mathrm{EVI} = 2.5 \times \frac{(\rho\mathrm{NIR}-\rho\mathrm{Red})}{(\rho\mathrm{NIR}+6\times\rho\mathrm{Red}-7.5\times\rho\mathrm{Blue}+1)}$	Huete et al. (2002)
EVI 2	$\mathrm{EVI2} = 2.5 \times \frac{(\rho\mathrm{NIR}-\rho\mathrm{Red})}{(\rho\mathrm{NIR}+2.4\times\rho\mathrm{Red}+1)}$	Jiang et al. (2008)

ρ=reflectance

Machine learning regression methods

Using *The Unscrambler*® (CAMO Software—Norway), two machine learning methods ($p \leq 0.05$) were used to correlate the Sentinel-2 spectral data (independent variables) to soybean yield (dependent variable):

Partial least square regression (PLSR)

PLSR is a multivariate regression method that performs the linear correlation between predictor variables, i.e., spectral bands, and dependent variable, i.e., soybean yield, by the determination of orthogonal base vectors, or latent variables, that account for most of the variation in the response variable, generating a linear model composed of waveband scaling coefficients to transform the spectral data (Yendrek et al., 2017). PLSR was developed based on the optimum number of latent variables, considering the lowest value of root mean square error (RMSE) and highest coefficient of determination (R^2) in the cross-validation procedure (Souza et al., 2013). For more information about PLSR as a machine learning regression method for remote sensed data, please refer to Maimaitijiang et al. (2020).

Support vector regression (SVR)

SVR is a non-parametric regression method that performs the non-linear correlation between predictor variables, i.e., spectral bands, and dependent variable, i.e., soybean yield, by fitting an optimal hyperplane (Ashourloo et al., 2016). A hyperplane can be defined as boundaries in high-dimensional space that classify the dependent variable for a regression task and, in the present research, was identified by using a non-linear kernel function (the Radial Basis Function) (Kayad et al., 2019; Kamir et al., 2020). In SVR, the number of parameters used in the model changes according to the input data and might be higher as the input data volume increases (Kayad et al., 2019).

Accuracy assessment

To assess the performance of PLSR and SVR models, a 10-fold cross-validation procedure was adopted, in which the input data is divided in ten subsets with roughly equal proportion and ten iterations, withholding one subset at a time. On each iteration, a training model is generated and then validated in the withheld subset. The overall accuracy of the model is obtained by the average of the results from each iteration (Guan et al., 2017, Hunt et al., 2019). Two statistical metrics were used to evaluate the model performance: the root mean squared error (RMSE), and R^2 (the proportion of variability in crop yield explained by the model). Hence, results from PLSR and SVR described in the present manuscript refer to the cross-validation procedure.

Sample selection

After all yield maps and Sentinel-2 images were processed, they were clipped to each field and pixels from field edges were excluded to avoid negative effects from the spectral mixing with adjacent areas (Brown et al., 2013). The remaining pixels had their pixel values (containing yield and reflectance from each band from each image) extracted using a 40 m $\times$ 40 m grid sampling, selecting, thus, from 20 to 30% of all pixels. Pixels extracted from the 40 m $\times$ 40 m grid sampling were used to conduct the statistical analysis described in section ‘Strategies for yield monitoring’, while yield and spectral response from all pixels were used to apply the spectral models for within-field soybean yield mapping, described in section ‘Mapping within field soybean yield’.

Strategies for yield monitoring

Yield monitoring through single Sentinel-2 images

Fields with similar sowing dates within each farm (Table 1) were inserted into the same dataset, i.e., within AV farm, fields were analyzed together; within TP farm, fields were analyzed together; and within MS farm, four groups of fields were analyzed, one containing MS 1, 2, 3 and 4, one containing only MS 5, one containing only MS 6, and one containing MS 7 and 8.

To assess the correlation between yield and Sentinel-2 images on each overpass date (Table 2), a linear regression was established between yield values and the derived vegetation indices (Table 3), and evaluated by the coefficient of determination (R^2). To check whether the use of the nine spectral bands from Sentinel-2 can deliver competitive results, reflectance values from all spectral bands from each overpass date (Table 2) were correlated to yield using PLSR and SVR models and evaluated by the coefficient of determination (R^2) from the cross-validation procedure.

Yield monitoring through Sentinel-2 images pooled across cropping season

To assess the contribution of using all available Sentinel-2 images across the cropping season (Table 2), each derived vegetation index (Table 3) had its temporal response pooled into the same dataset and correlated to yield using PLSR and SVR. Under the same perspective, the nine spectral bands from all available Sentinel-2 images were pooled into the same dataset and correlated to yield using PLSR and SVR. Additionally, all VIs and all spectral bands from all available Sentinel-2 images were pooled into the same dataset to evaluate whether combining: (1) all vegetation indices from all available images; or (2) all vegetation indices and all spectral bands from all available images; could deliver competitive results for yield prediction. The accuracy was assessed by the R^2 and RMSE from the cross-validation procedure. Fields with similar sowing dates within each farm, as described in section 'Yield monitoring through single Sentinel-2 images', were analyzed together.

Yield monitoring through Sentinel-2 images at the R5 phenological stage

To make all fields from all farms comparable among them, despite their differences in sowing dates, spectral data acquired at the R5 phenological stage (which corresponds to the grain filling stage—Fehr & Caviness, 1977) were used. The R5 phenological stage has been recognized to have the highest correlation with grain yield using soybean spectral response (Bolton & Friedl, 2013; Carneiro et al., 2020; Kross et al., 2020 Sakamoto 2020, Crusiol et al., 2021a).

Considering that soybean phenology might be influenced by environmental conditions and the cultivar used, affecting the time of occurrence of the R5 phenological stage, images from 80 to 100 days after sowing (DAS), for each field, were considered as correspondent to the R5 phenological stage. The selection of this time interval was based on research findings addressing the R5 phenological stage for soybean spectral monitoring and based on results from field experiments addressing the phenology characterization of multiple soybean cultivars at the National Soybean Research Center (Embrapa Soja—Brazilian Agricul-

ture Research Corporation—located in Londrina municipality, less than 75 km far from the study areas). Thus, at this stage, the available Sentinel-2 images for each field were merged for each spectral band separately, promoting the standardization of the number of input data (one image containing nine spectral bands) for crop yield monitoring at the R5 phenological stage, as suggested by Guan et al. (2017) and Gómez et al. (2019).

PLSR and SVR models (correlating yield and Sentinel-2 spectral bands) were developed, separately, for each field, where 15 field-based models were generated (one for each of them). Besides that, PLSR and SVR models were developed, separately, for each farm, by aggregating within each farm, the pixels used to develop the field-based models. As a result, 3 farm-based models were generated (one for each farm).

Aiming at developing a global-based model, yield and spectral data from all fields from all farms were analyzed together, by aggregating the pixels used to develop all the 15 field-based models. At this stage, yield values were correlated to: the derived vegetation indices (Table 3) and evaluated by the coefficient of determination (R^2), and to Sentinel-2 spectral bands, using PLSR and SVR models, and evaluated by the R^2 and RMSE from the cross-validation procedure.

The input data for each model (field-, farm- and global-based) consisted of one image (obtained after merging the available images between 80 and 100 DAS—Table 2) containing the nine Sentinel-2 spectral bands; and the sampling size for each field, extracted from the 40 m × 40 m grid sampling (section 'Sample selection'), varied from 20 to 30% of all pixels.

To provide further evidence of the predictive capacity for soybean yield of the developed global-based models, data were further analyzed by splitting into calibration and validation subsets (containing 75% and 25% of data respectively). Thus, the calibration subset was used to generate an additional prediction model, whose accuracy was assessed by R^2 and RMSE from the cross-validation procedure, while the remaining 25% was used as validation subset, whose accuracy was assessed by the R^2 and RMSE from the linear correlation between observed and predicted yield.

Comparing satellite-based and ground-based data for yield monitoring

Aiming at comparing spectral data obtained through Sentinel-2 images and through field campaigns, a field experiment was undertaken at the National Soybean Research Center (Embrapa Soja), a branch of the Brazilian Agricultural Research Corporation, located in Londrina Municipality, Paraná State, Southern Brazil (Fig. 1), in the 2016/2017, 2017/2018 and 2018/2019 cropping seasons. The motivation of comparing the results obtained from Sentinel-2 images to results obtained at leaf scale, on field experiments, was to provide further evidence of the relation between soybean yield and spectral response at both levels of data acquisition, strengthening the soybean yield monitoring ability.

The spectral response of ten soybean genotypes and cultivars, with different responses to water availability and submitted to non-irrigated (rainfed) condition, meeting, as close as possible, the natural rainfed conditions found in the three evaluated farms, was assessed at the R5 phenological stage. Leaf-based reflectance was collected at 89, 96 and 94 DAS (2016/2017, 2017/2018 and 2018/2019 cropping seasons respectively) on the central leaflet of the fullest expanded third trifoliate leaf from the top using the FieldSpec 3 Jr spectroradiometer (Analytical Spectral Devices, Boulder, CO, USA), with a spectral resolution of 3 nm

between 350 and 1,400 nm and 30 nm between 1,400 and 2,500 nm. The plant probe device, with an internal 99% reflectance board (Spectralon®—used as reflectance standard), and a 1% reflectance opaque and black board (used during the spectral assessment to ensure pure leaf reflectance spectra collection) was used to prevent illumination interferences of adjacent targets and atmospheric scattering and attenuation. On each plot, four leaf-based spectral subsamples were collected and then averaged, resulting in the samples used for data analysis, in a total of 224 subsamples and 56 spectral samples from the three cropping seasons.

The hyperspectral data were then resampled to meet the Sentinel-2 spectral bands, named from now on as Sentinel-2-like. A linear regression was stablished between yield and the derived vegetation indices (Table 3), evaluated by the coefficient of determination (R^2). To assess the correlation between yield and Sentinel-2-like spectral bands, PLSR and SVR models were generated and evaluated by the R^2 and RMSE from the cross-validation procedure. Following the procedure adopted in section 'Yield monitoring through Sentinel-2 images at the R5 phenological stage', to provide further evidence of the predictive capacity for soybean yield of the developed Sentinel-2-like models, data were further analyzed by splitting into a calibration subset (containing 75% of data), whose accuracy was assessed by the R^2 and RMSE from the cross-validation procedure, and the remaining 25% of data were used as validation subset, whose accuracy was assessed by the R^2 and RMSE from the linear correlation between observed and predicted yield.

Mapping within-field soybean yield

Aiming at mapping within-field soybean yield, the field-based, farm-based and global-based SVR models, generated in section 'Yield monitoring through Sentinel-2 images at the R5 phenological stage' using a 40 m × 40 m grid sampling (section 'Sample selection'), were applied to all pixels from each field. To assess the accuracy of yield mapping, using the developed SVR models applied to all pixels, a linear regression was established between predicted yield (mapped) and observed yield (recorded by a combine harvester with yield monitor) and the R^2 and RMSE were used to evaluate the trend between both datasets (predicted and observed).

Extra SVR models were developed, for each field, using Sentinel-2 and yield values from all pixels, to ensure the unbiased selection of yield and spectral samples from the 40 m × 40 m grid sampling. Their accuracy was assessed by the R^2 and RMSE from the cross-validation procedure.

Results

Characterization of soybean yield on each evaluated farm

Figure 2 displays the boxplot for soybean yield measured from all pixels and from the 40 m × 40 m grid sampling (section 'Sample selection') on each farm. Yield values extracted from the 40 m × 40 m grid sampling had similar distribution compared to yield values extracted from all pixels within each farm. This similar distribution denotes the effectiveness in selecting from 20 to 30% of pixels from each farm, representing the variability of soybean yield.

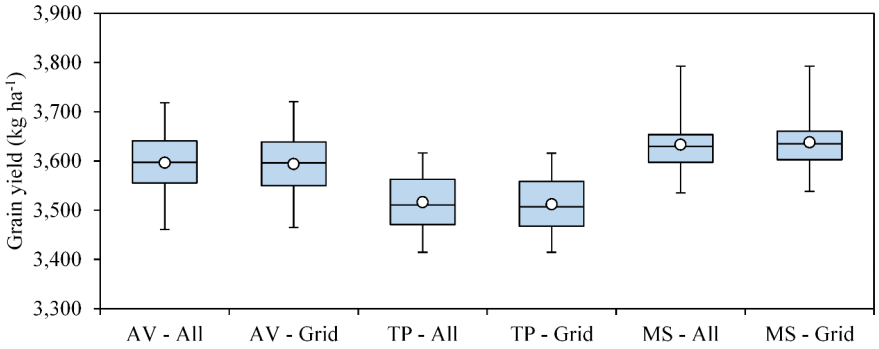


Fig. 2 Boxplot of soybean grain yield on Água Viva (AV), Tupinambá (TP) and Mauá da Serra (MS) farms from all pixels (All) and from the 40 m × 40 m grid sampling (Grid)

Relation between yield and single Sentinel-2 images

Tables 4 and 5 present the coefficients of determination (R^2) for soybean yield using vegetation indices and from PLSR and SVR using Sentinel-2 spectral bands on AV, TP and MS farms for each single image across cropping season. For all evaluated farms (on MS farm, four groups of fields were analyzed according to their sowing date, as described in section 'Yield monitoring through single Sentinel-2 images'), a temporal pattern regarding the most

Table 4 Coefficients of determination (R^2) for soybean yield using vegetation indices and from PLSR and SVR using Sentinel-2 spectral bands on AV and TP farms

		Nov.	Dec.	Jan.		Feb.
		29	29	18	28	12
AV	BNDVI	0.05	0.01	0.01	0.01	0.01
	EVI2	0.05	0.01	0.03	0.01	0.01
	EVI	0.04	0.01	0.06	0.14	0.01
	GNDVI	0.05	0.01	0.03	0.02	0.11
	NDVI	0.05	0.01	0.02	0.01	0.01
	NDRE	0.06	0.01	0.07	0.04	0.01
	NDII	0.05	0.01	0.34	0.14	0.01
	NDII2	0.04	0.01	0.42	0.21	0.02
	PLSR[a]	0.70	0.59	0.70	0.53	0.60
	SVR[a]	0.82	0.68	0.80	0.55	0.62
TP	BNDVI		0.01	0.08	0.17	0.18
	EVI2		0.01	0.05	0.14	0.16
	EVI		0.04	0.01	0.01	0.11
	GNDVI		0.01	0.01	0.18	0.28
	NDVI		0.01	0.05	0.14	0.16
	NDRE		0.01	0.01	0.18	0.29
	NDII		0.18	0.41	0.01	0.07
	NDII2		0.09	0.26	0.04	0.01
	PLSR[a]		0.34	0.42	0.53	0.57
	SVR[a]		0.40	0.54	0.60	0.61

Table 5 Coefficients of determination (R^2) for soybean yield using vegetation indices and from PLSR and SVR using Sentinel-2 spectral bands on MS farm

		Nov.	Dec.	Jan.		Feb.		Mar.		
		29	29	18	28	12	27	03	08	13
MS 1–4	BNDVI	0.05	0.01	0.02	0.03	0.01				
	EVI2	0.05	0.01	0.02	0.01	0.01				
	EVI	0.05	0.01	0.01	0.02	0.01				
	GNDVI	0.07	0.01	0.08	0.04	0.01				
	NDVI	0.05	0.01	0.02	0.01	0.01				
	NDRE	0.08	0.01	0.01	0.01	0.01				
	NDII	0.01	0.01	0.01	0.12	0.02				
	NDII2	0.01	0.01	0.05	0.17	0.01				
	PLSR[a]	0.55	0.44	0.29	0.26	0.39				
	SVR[a]	0.54	0.48	0.35	0.27	0.35				
MS 5	BNDVI		0.13	0.01	0.13	0.02	0.22	0.18		
	EVI2		0.02	0.01	0.18	0.07	0.26	0.18		
	EVI		0.08	0.08	0.06	0.07	0.27	0.18		
	GNDVI		0.09	0.06	0.09	0.07	0.23	0.10		
	NDVI		0.02	0.01	018	0.07	0.26	0.19		
	NDRE		0.02	0.01	0.15	0.12	0.25	0.10		
	NDII		0.21	0.05	0.02	0.02	0.29	0.28		
	NDII2		0.15	0.08	0.08	0.08	0.29	0.27		
	PLSR[a]		0.50	0.26	0.33	0.26	0.29	0.20		
	SVR[a]		0.52	0.22	0.36	0.25	0.28	0.23		
MS 6	BNDVI			0.04	0.01	0.05	0.08	0.06	0.01	0.46
	EVI2			0.05	0.01	0.05	0.32	0.18	0.06	0.38
	EVI			0.05	0.01	0.01	0.09	0.03	0.03	0.35
	GNDVI			0.03	0.12	0.01	0.02	0.02	0.09	0.42
	NDVI			0.05	0.01	0.05	0.32	0.18	0.06	0.38
	NDRE			0.01	0.05	0.01	0.11	0.01	0.07	0.36
	NDII			0.01	0.01	0.48	0.71	0.62	0.59	0.40
	NDII2			0.01	0.01	0.42	0.69	0.53	0.56	0.36
	PLSR[a]			0.27	0.29	0.54	0.76	0.70	0.70	0.69
	SVR[a]			0.32	0.29	0.57	0.78	0.70	0.71	0.75
MS 7–8	BNDVI			0.25	0.16	0.08	0.01	0.01	0.01	0.06
	EVI2			0.29	0.18	0.06	0.04	0.02	0.01	0.06
	EVI			0.31	0.17	0.03	0.07	0.04	0.02	0.05
	GNDVI			0.24	0.24	0.17	0.01	0.01	0.07	0.14
	NDVI			0.29	0.18	0.06	0.04	0.02	0.01	0.05
	NDRE			0.27	0.25	0.17	0.01	0.01	0.07	0.15
	NDII			0.23	0.22	0.14	0.02	0.08	0.03	0.03
	NDII2			0.26	0.23	0.12	0.03	0.08	0.03	0.02
	PLSR[a]			0.27	0.24	0.22	0.18	0.28	0.31	0.40
	SVR[a]			0.29	0.26	0.21	0.26	0.35	0.36	0.44

feasible time for yield monitoring was not observed either using vegetation indices or Sentinel-2 spectral bands. For some fields, the highest correlation was observed at the beginning of cropping season (e.g., AV, MS 1–4 and MS 5), while some fields presented the highest

correlation at the middle (e.g., MS 6) and at the ending of the cropping season (e.g., TP and MS 7–8).

Besides the absence of temporal pattern, the correlation between vegetation indices and yield presented, for each VI, both positive and negative correlations on different images, and non-significant correlations were frequently observed on the three farms, precluding the identification of the most feasible vegetation index for within-field soybean yield monitoring.

Among the evaluated vegetation indices, those derived from near-infrared and shortwave infrared bands (NDII and NDII2) demonstrated outstanding results, achieving competitive results with the PLSR and SVR models and even outperforming them at MS 5 (Table 5) on March 3rd. However, using machine learning regression methods demonstrated larger stability in the trend between spectral response and soybean yield. Hence, the use of all Sentinel-2 bands through PLSR and SVR demonstrated, for all fields and dates across cropping season (with the exception of MS 5 on March 3rd), higher accuracy for within-field yield prediction compared to the use of single VIs.

Relation between yield and Sentinel-2 images pooled across cropping season

Figures 3 and 4 present the results derived from PLSR and SVR using Sentinel-2 images pooled across the cropping season on AV, TP and MS farms. At this stage, each vegetation index (Table 3) had their temporal response correlated to yield using PLSR and SVR.

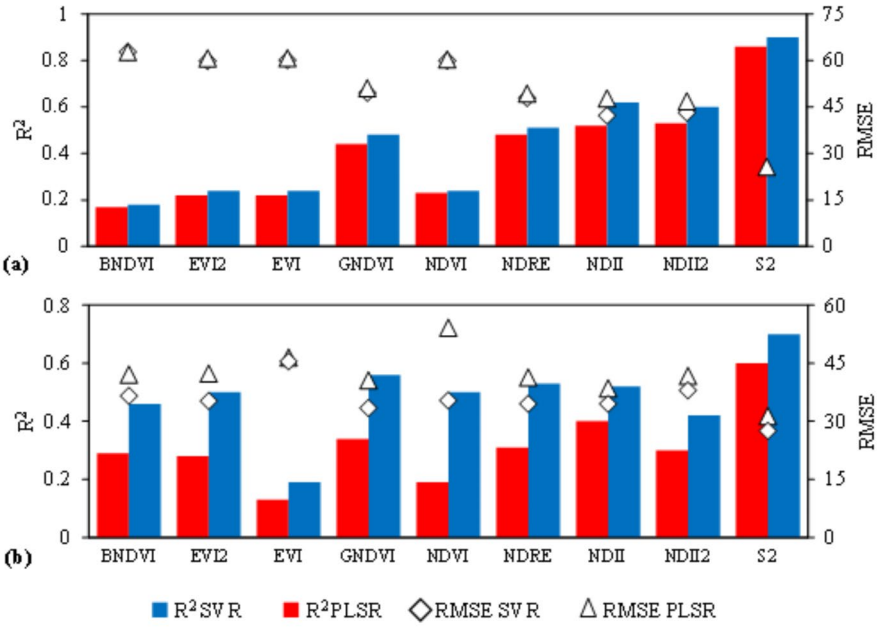


Fig. 3 R^2 and RMSE derived from PLSR and SVR for soybean yield prediction using pooled vegetation indices and Sentinel-2 spectral bands (S2) across cropping season on AV (a) and TP (b) farms

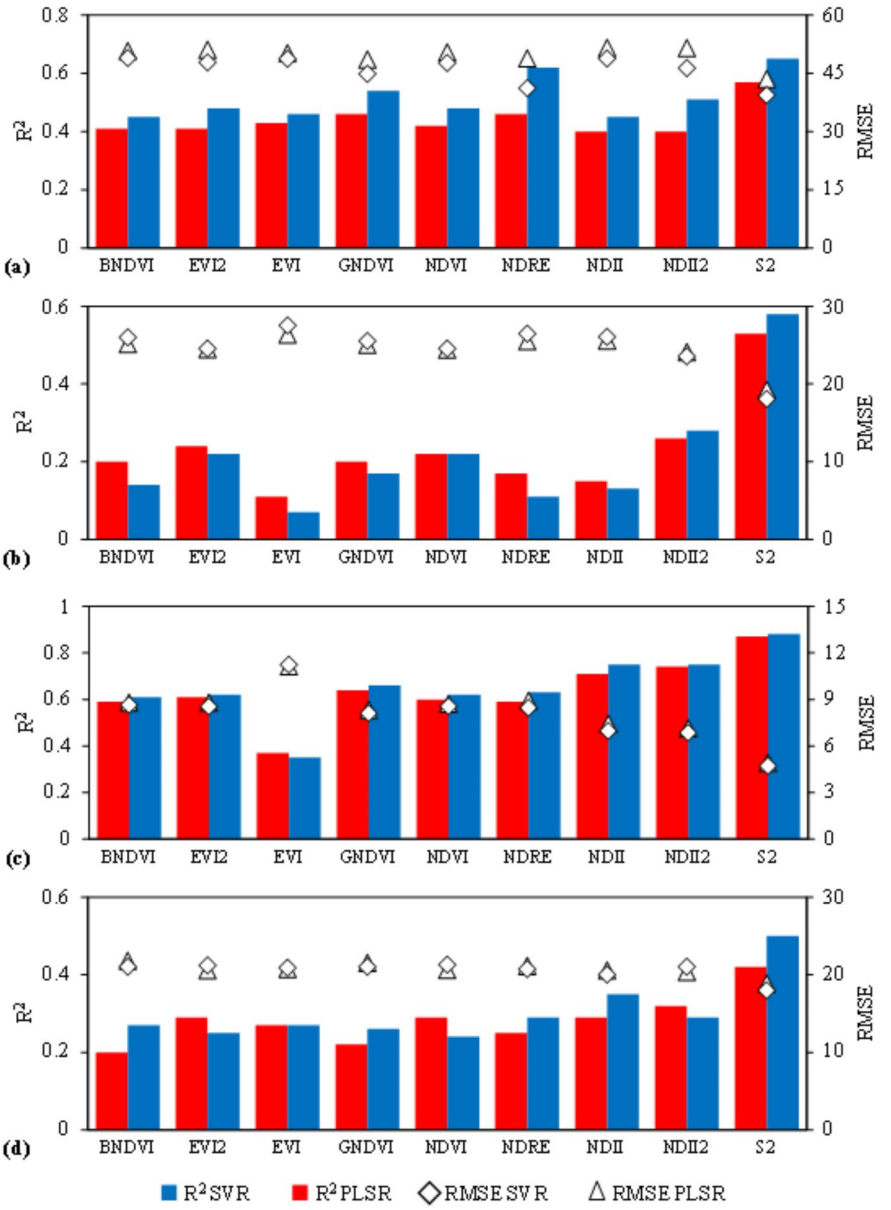


Fig. 4 R^2 and RMSE derived from PLSR and SVR for soybean yield prediction using pooled vegetation indices and Sentinel-2 spectral bands across cropping season on MS farm: MS 1–4 (a), MS 5 (b), MS 6 (c) and MS 7–8 (d)

Similarly, the nine spectral bands from all available images were also pooled into the same dataset and correlated to yield using PLSR and SVR.

For all vegetation indices, the use of their temporal response pooled into the same dataset and correlated to yield using PLSR and SVR demonstrated higher correlation compared to the use of vegetation index from a single image (Tables 4 and 5). As has been observed using single images, the use of pooled data across cropping season did not demonstrate a pattern regarding the most feasible VI for within-field soybean yield monitoring.

Although the accuracy for within-field yield prediction had been increased by the use of vegetation index pooled through the cropping season, the use of the nine spectral bands from all available Sentinel-2 images across the cropping season demonstrated the highest correlation to yield compared to information acquired at single image (Tables 4 and 5) and temporal response from VIs.

To evaluate whether combining: (1) all vegetation indices from all available images; or (2) all vegetation indices and all spectral bands from all available Sentinel-2 images; could deliver competitive results for yield prediction, PLSR and SVR models were developed for each farm and are presented in Fig. 5.

The use of all vegetation indices (Table 3) from all Sentinel-2 images (Table 2) did not result in higher accuracy compared to the use of all spectral bands from all Sentinel-2 images across cropping season. Besides that, adding all VIs to all spectral bands from all images demonstrated a meaningless contribution to both PLSR and SVR models for yield prediction.

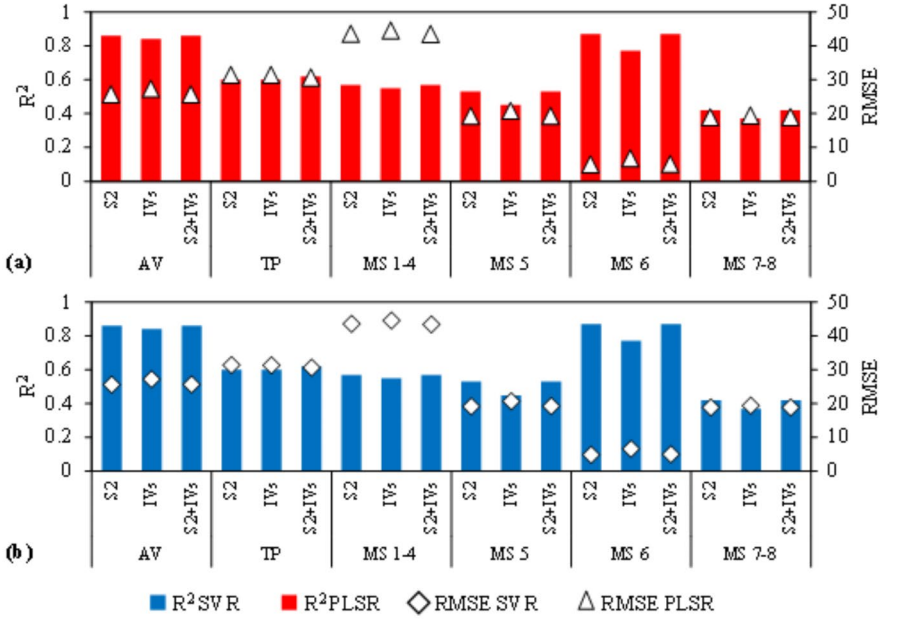


Fig. 5 R^2 and RMSE derived from PLSR (a) and SVR (b) regressions for soybean yield prediction using pooled Sentinel-2 spectral bands (S2), vegetation indices (VIs), and Sentinel-2 spectral bands + vegetation indices (S2+VIs) from all images across cropping season

Relation between yield and Sentinel-2 images at the R5 phenological stage

The results derived from PLSR and SVR using Sentinel-2 images at the R5 phenological stage for each field (field-based models) and each farm (farm-based models), presented in Table 6, demonstrated that the accuracy in soybean yield prediction presented different results for different production areas. Regarding the global-based model, Fig. 6 (a) displays the boxplot for soybean yield measured from all fields from all farms analyzed together (global-based model) and its splitting into calibration (75%) and validation (25%) subsets. The boxplot for soybean yield measured at Embrapa Soja is presented in Fig. 6 (b). The

Table 6 R^2 and RMSE derived from PLSR and SVR regressions for soybean yield using Sentinel-2 spectral bands at the R5 phenological stage on each field (field-based models) and each farm (farm-based models)

		Yield (kg ha^{-1})			PLSR		SVR	
Farm	Field	Mean	σ	Model-based	R^2	RMSE	R^2	RMSE
AV	AV 1	3669.96	28.38	Field	0.78	13.08	0.79	12.85
	AV 2	3597.75	12.49	Field	0.17	11.35	0.29	10.52
	AV 3	3526.75	38.30	Field	0.49	27.03	0.53	26.20
	All	3594.27	65.04	Farm	0.63	39.36	0.70	35.77
TP	TP 1	3575.90	11.47	Field	0.21	10.42	0.07	11.06
	TP 2	3567.39	25.43	Field	0.37	20.35	0.44	19.43
	TP 3	3479.87	43.34	Field	0.30	36.27	0.26	37.32
	TP 4	3490.54	23.64	Field	0.46	17.40	0.48	17.03
	All	3512.18	50.04	Farm	0.52	34.47	0.60	31.47
MS	MS 1	3653.96	30.26	Field	0.06	29.99	0.28	25.82
	MS 2	3598.22	25.06	Field	0.73	13.00	0.54	17.04
	MS 3	3710.81	42.04	Field	0.26	36.40	0.38	33.82
	MS 4	3750.40	25.72	Field	0.16	23.63	0.18	23.63
	MS 5	3577.43	27.74	Field	0.25	24.24	0.32	23.08
	MS 6	3621.88	13.67	Field	0.71	7.21	0.72	7.24
	MS 7	3642.59	8.71	Field	0.13	8.22	0.23	7.49
	MS 8	3644.05	25.86	Field	0.28	21.83	0.34	21.04
	All	3638.30	56.93	Farm	0.52	39.06	0.60	35.86

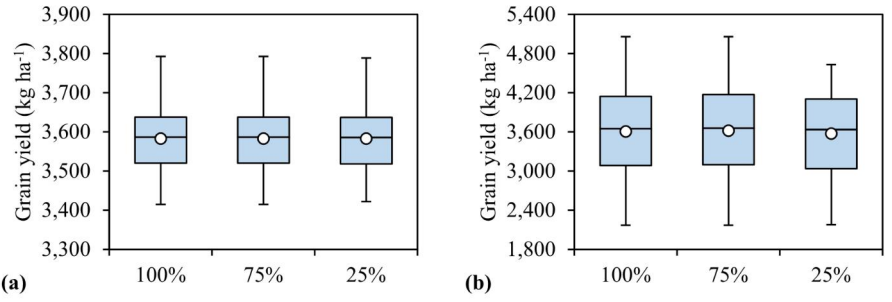


Fig. 6 Boxplot of soybean grain yield from Global-based model (a) and from Embrapa Soja (b), comprising 100%, 75% and 25% of data

portioning of the global-based model and data from Embrapa Soja into 75% and 25% was undertaken to provide further evidence of their ability for soybean yield prediction. Hence, 75% of data was used to generate an additional prediction model, while the remaining 25% was used to perform its validation. It is possible to verify, within each dataset, a similar

Table 7 Coefficients of determination (R^2) for soybean yield using vegetation indices derived from Sentinel-2 images (Sentinel-2) and ground-based Sentinel-2-like spectral data (Embrapa Soja) at the R5 phenological stage

	Sentinel-2	Sentinel-2-like
BNDVI	0.13	0.10
EVI2	0.04	0.15
EVI	0.12	0.17
GNDVI	0.03	0.01
NDVI	0.04	0.16
NDRE	0.01	0.06
NDII	0.21	0.01
NDII2	0.15	0.01

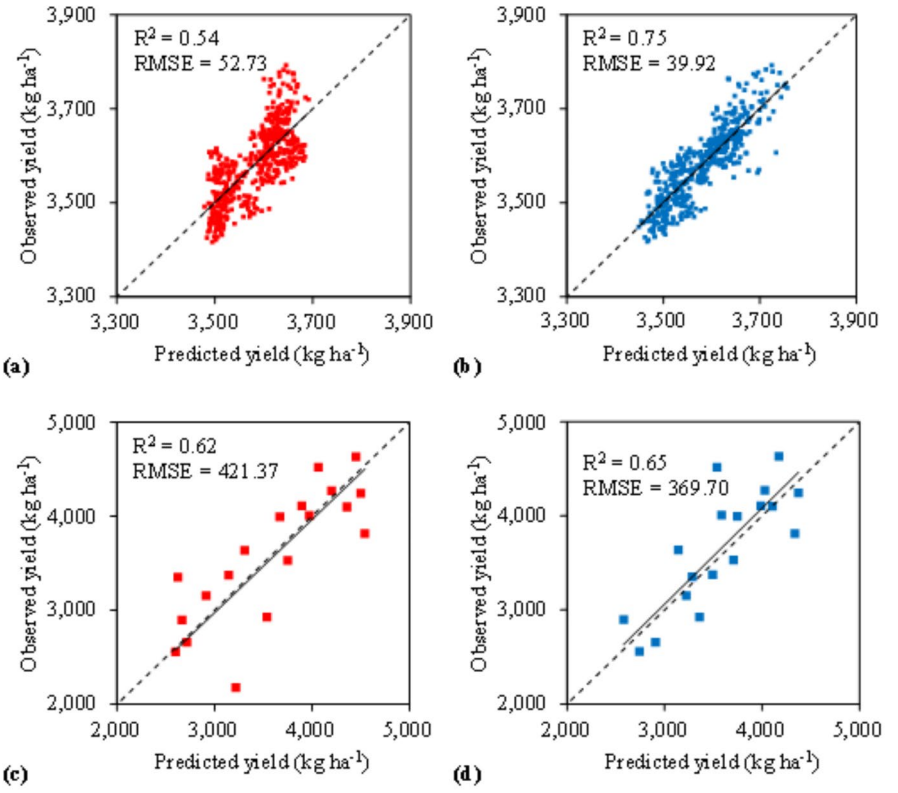


Fig. 7 Scatterplot between observed and predicted (external validation) soybean yield using Sentinel-2 spectral bands through PLSR (a) and SVR (b) and using ground-based Sentinel-2-like spectral data (Embrapa Soja) through PLSR (c) and SVR (d)

distribution among 100% of data (global-based model and Embrapa Soja) and their splitting into 75% and 25% subsets.

The average spectral response of soybean crop at the R5 phenological stage from the global-based model (Sentinel-2) and from Embrapa Soja (Sentinel-2-like), presented in a supplementary figure, demonstrated a similar response across the visible and shortwave-infrared spectrum with larger differences, however, at bands 7 and 8a (780 nm and 860 nm respectively). Table 7 presents the coefficients of determination (R^2) for soybean yield using vegetation indices at R5 the phenological stage derived from Sentinel-2 (global-based model) and Sentinel-2-like (Embrapa Soja) data. Table 8 presents the results derived from PLSR and SVR using, at the R5 phenological stage, Sentinel-2 images (global-based model) and Sentinel-2-like (Embrapa Soja) and their 75% subsets.

As has been observed, either for single images or images pooled across cropping season, the use of vegetation indices for monitoring within-field soybean yield demonstrated low values of R^2, both using satellite and ground-based data (Table 7). On the other hand, the use of visible, near-infrared and shortwave infrared spectral bands (Table 8) demonstrated higher correlation with within-field soybean yield, with R^2 equal to 0.75 for global-based model and 0.63 for Sentinel-2-like (Embrapa Soja) model, using SVR approach. Besides that, the generation of an additional prediction model, using 75% of data, demonstrated similar accuracy compared to the global-based and Sentinel-2-like models. These two additional models (generated using 75% of data from global-based and Embrapa Soja) were applied to the remaining 25% of data, and Fig. 7 presents the results from the external validation.

It is possible to observe a similar trend between observed and predicted soybean yield (using Vis/NIR/SWIR spectral bands) either from satellite or ground-based remote sensed data. Those findings are important to support future research addressing soybean yield monitoring using spectral data collected from multiple platforms and integrated among them. The results from SVR from Sentinel-2 spectral bands and Sentinel-2-like (Embrapa Soja) revealed higher accuracy in soybean yield prediction compared to PLSR. Although for Sentinel-2-like (Embrapa Soja), results obtained for SVR were slightly higher than PLSR, with R^2 equal to 0.65 and 0.62 respectively, for Sentinel-2 (global-based model) larger differences in their accuracies were observed (R^2 equal to 0.75 and 0.54 respectively). Despite the differences in the accuracy from both regression methods, the strong and positive relation between observed and predicted soybean yield is emphasised. The regression analysis with an intersection passing through the origin (y = bx) revealed an adjusted model (y = 1.0002x and y = 1.0004x) for Sentinel-2 (global-based model) using PLSR and SVR respectively;

Table 8 R^2 and RMSE derived from PLSR and SVR for soybean yield at the R5 phenological stage using Sentinel-2 images and ground-based Sentinel-2-like spectral data (Embrapa Soja) comprising 100% and 75% of the spectral dataset

		PLSR		SVR	
Dataset	Model	R^2	RMSE	R^2	RMSE
Sentinel-2	Global-based 100%	0.56	51.76	0.75	38.82
	Global-based 75%	0.55	52.16	0.75	39.16
Sentinel-2-like	Embrapa Soja 100%	0.63	409.58	0.63	404.20
	Embrapa Soja 75%	0.61	409.79	0.65	393.88

and (y = 0.9910x and y = 1.0206x) for Sentinel-2-like (Embrapa Soja) using PLSR and SVR respectively, indicating a similar trend in the adjusted models.

Mapping within-field soybean yield

Considering the better performance of SVR in relation to PLSR for soybean within-field monitoring, SVR models from the field, farm and global-based models were used for mapping within field soybean yield. Figure 8 presents the R^2 and RMSE between observed and predicted yield from SVR models at the R5 phenological stage using the global-based model (developed using information from all fields from all farms), farm-based models (developed using information within each farm) and field-based models (developed using information from each field separately). The aforementioned models were developed using data extracted from a 40 m × 40 m grid sampling and were applied to all pixels to perform

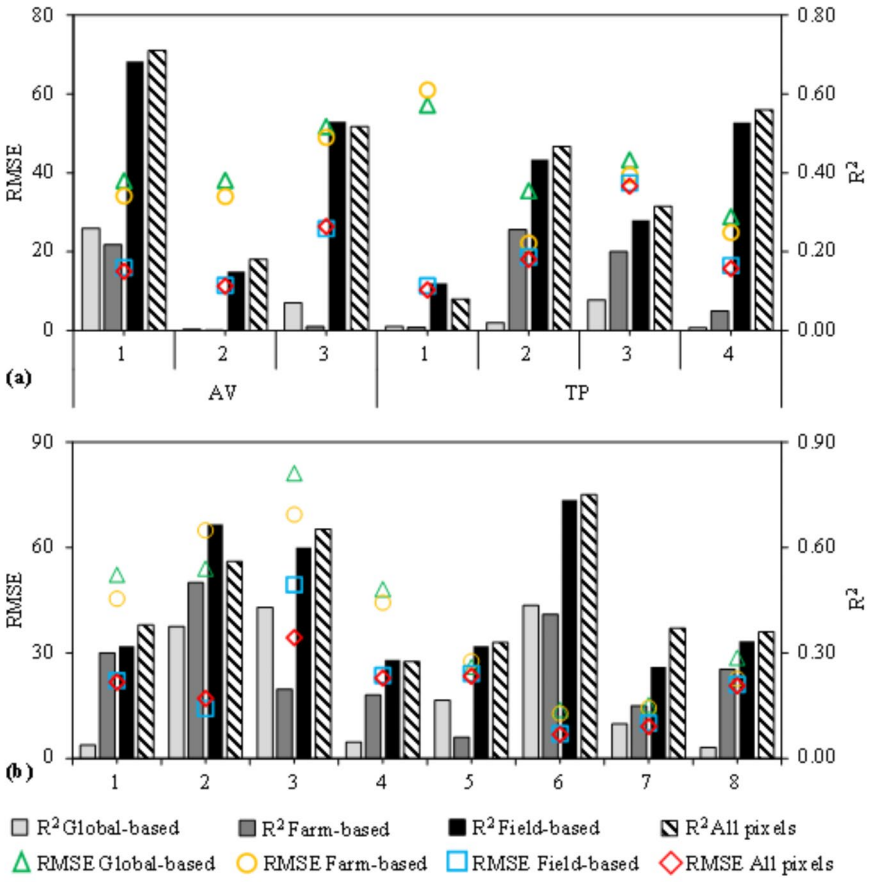


Fig. 8 R^2 and RMSE between observed and estimated soybean yield on AV and TP farms (a) and MS farm (b) using Sentinel-2 spectral bands through SVR applying the Field-based, Farm-based and Global-based models and the model generated separately for each field, using all pixels

the yield mapping. To ensure the unbiased selection of samples from the grid sampling, extra SVR models were developed, for each field, using values from all pixels, labelled in Fig. 8 as 'all pixels' (as described in section 'Mapping within field soybean yield').

For all the 15 fields, the use of field-based models delivered the highest accuracy for within-field yield mapping. It was observed that the accuracy for yield prediction increased from the most complex dataset (global-based model) to models that take into account the complexity from each farm (farm-based model), reaching the highest accuracy using field-based models. The observed and predicted yield maps from each field using the respective SVR field-based model are presented in Fig. 9. For each field, the upper and lower yield limits were set to the same value for observed and predicted maps, allowing their visual comparison.

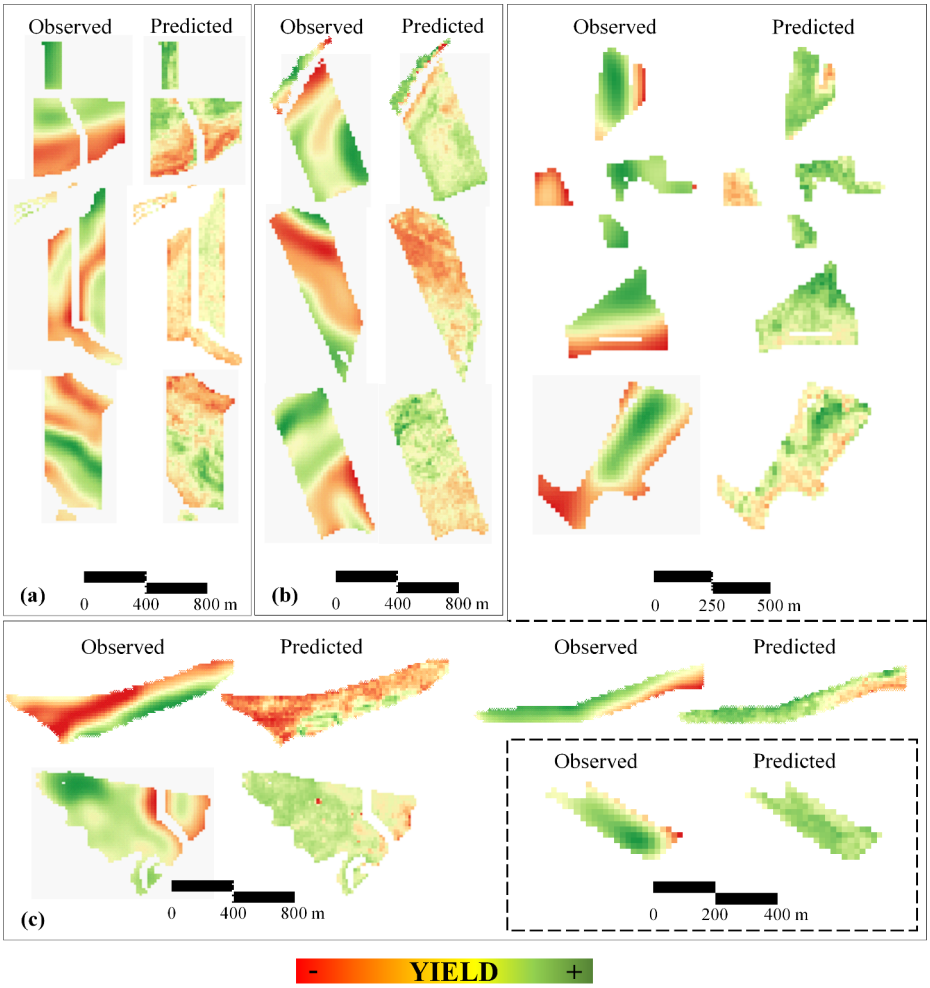


Fig. 9 Observed and predicted yield maps from AV (a), TP (b) and MS (c) farms

The comparison between the field-based models (Fig. 8, black bars) and the extra SVR models (developed, for each field, using values from all pixels—Fig. 8, white bars with black stripes) demonstrated competitive results in all the 15 evaluated fields, suggesting that a small number of samples (from 20 to 30% in each field, selected from the 40 m × 40 m grid sampling) can be used for yield mapping without loss of accuracy. Performing the regression analysis between the coefficient of determination and root mean square error obtained by the field-based models and the extra SVR models, a strong positive correlation was observed. The regression analysis with an intersection passing through the origin (y = bx) revealed an adjusted model ($y = 1.0266x$ and $y = 0.8963x$) with coefficients of determination (R^2) equal to 0.989 and 0.977 for their R^2 and RMSE respectively.

Discussion

The yield dataset used in the present manuscript demonstrated variability not only within farm, but also between them (Fig. 2). According to Embrapa Soja (2013), the soybean crop is deeply affected by, among others, weather conditions, soil fertility and management, seed technology and cultivar, invasive plants, disease and insect management, and technical adjustments during the establishment of the crop, such as sowing date and plant population. Although the farms evaluated are located within the same geographical zone (Fig. 1), it was expected that the interferences aforementioned had led to differences in soybean growth conditions and, consequently, differences in yield among and within them.

Relation between yield and single Sentinel-2 images

The time interval for optimal within-field yield monitoring demonstrated to be different on each farm and among the evaluated fields (Tables 4 and 5). The characterization of the most feasible time for yield prediction across cropping season might be difficult due to the complex relationships between crop growth and the biotic and abiotic factors (Kross et al., 2020), such as weather conditions and soil and crop management, which may not be the same in all production areas (Robson et al., 2017). According to Esquerdo et al. (2011), there is a trade-off between early prediction and accuracy in soybean yield prediction, suggesting that the accuracy in yield prediction increases toward the end of the season (Gómez et al., 2019). Al-Gaadi et al. (2016), reported limitations in the detection of the best time across cropping season for potato yield prediction.

Besides the lack in temporal pattern, the evaluated vegetation indices did not demonstrate a trend with yield, presenting, in the different images, both positive and negative correlations for the same index, which also precluded the identification of the most feasible vegetation index for within-field soybean yield monitoring. Leon et al. (2003), obtained, for the same VI, both positive and negative correlations between soybean yield and vegetation indices without temporal pattern on three fields. Using VIs, Ali et al. (2019) reported different accuracies for yield prediction in several crop types according to the phenological stage. Segarra et al. (2020), predicting wheat yield at regional level, and Suarez et al. (2020), predicting carrot yield, reported limitations in the identification of the most accurate VI.

When developing yield prediction models using all Sentinel-2 spectral bands and machine learning regression methods, higher accuracy compared to the use of single VIs

was observed, for all fields and dates across the cropping season (with the exception of MS 5 on March 3rd, when slightly higher R^2 was observed for NDII and NDII2),. The contribution of PLSR and SVR relies on the fact that multiple Vis-NIR-SWIR spectral bands were analyzed into the same model, enabling better characterization of the conditions of crop development across different wavelengths. The higher accuracy of spectral bands for yield prediction using machine learning regression methods have been reported in corn (Kayad et al. 2019), and soybean (Leon et al. 2003; Dado et al. 2020).

Relation between yield and Sentinel-2 images pooled across cropping season

It was demonstrated in Figs. 3 and 4 that pooling spectral information across the cropping season might contribute to enhance the accuracy of within-field yield prediction. This trend could be observed for all vegetation indices and also using all spectral bands from all available images. According to Gao et al. (2018), the use of combined spectral information through the cropping season contributes to better understanding the physiological dynamics of crops, better characterizing, therefore, their response to yield across time. Although many different biotic and abiotic factors can influence soybean yield, each of them has a more favorable time of occurrence across the cropping season. By this reasoning, using images pooled through time allows the acquisition of spectral information on different phenological stages and under the influence of their driving factors, which results in more accurate models for yield monitoring.

The use of pooled spectral bands from images across the cropping season for yield prediction has been addressed in soybean (Dado et al., 2020). Predicting soybean yield through NDVI images, Mercante et al. (2010) reported that pooled information from different dates delivered higher accuracy than a vegetation index from single images. As demonstrated in the analysis of single images, the contribution of PLSR and SVR relies on the fact that multiple Vis-NIR-SWIR spectral bands could better describe the influence of different stresses in crop yield across multiple wavelengths. Besides that, several images (from the early stages of crop development to the maturity stages) were analyzed into the same model, enabling, thus, the characterization of the biotic and abiotic factors that can influence yield values at different time across the cropping season, according to their timing of occurrence.

As demonstrated in Fig. 5, pooling all VIs from all available images presented similar accuracy compared to all spectral bands; and their combined use (adding all VIs to all spectral bands) did not contribute to enhance the model accuracy, which might be related to the fact that vegetation indices are calculated from spectral band reflectance and might, therefore, present similar information. Shoko and Mutanga (2017) reported that Sentinel-2 bands outperformed the use of vegetation indices or their combined use for the classification of grassland species. Hunt et al. (2019) demonstrated that spectral bands have higher correlation to wheat yield compared to vegetation indices and that their combined analysis did not enhance the prediction accuracy.

Relation between yield and Sentinel-2 images at the R5 phenological stage

Assessing the relation between yield and Sentinel-2 images at the R5 phenological stage, the accuracy in soybean yield prediction presented different results for different production areas for PLSR and SVR (Table 6), demonstrating that the influence of biotic and

abiotic factors on crop yield varies spatially and temporally between the assessed fields. Robson et al. (2017) and Rahman et al. (2018) demonstrated that the relation between spectral response and avocado yield changes from one cropping season to another and from field to field. Kross et al. (2020), monitoring within-field soybean yield, reported limitations due to the driving factors that affect each field, emphasizing the complex relation between crop development and its spectral response.

The accuracy in yield prediction using merged images at the R5 phenological stages was lower than using all available images (Figs. 3 and 4). However, the use of a standardized number of input data at the target period (one image containing nine spectral bands), as suggested by Guan et al. (2017) and Gómez et al. (2019), contributed to overcoming the limitation in the acquisition of images, mainly free of cloud cover, at similar periods across the cropping season for different fields, making the models developed more flexible for application and spatially extrapolated to new soybean areas.

Aiming at providing further evidence of the relation between soybean yield and spectral response using both satellite and ground-based data, the average spectral response from the global-based model was assessed and compared to the average spectral response from Embrapa Soja (Sentinel-2-like) dataset. A similar spectral response was found between both sensors, albeit with larger reflectance for the orbital sensor at near-infrared spectrum (supplementary figure). Lamquin et al. (2019) reported the closely related behavior of ground- and satellite-based sensors over a variety of calibration targets. Similar trends in the spectral response from Sentinel-2 images and simulated spectral bands from a hyperspectral sensor, albeit with differences in their reflectance intensities, have been reported by Toming et al. (2016) and Munyati et al. (2020). Although data from the global-based model and Embrapa Soja were not collected in the same area or time, their similar behavior, collected both at canopy level (Sentinel-2 images) and leaf level (Sentinel-2-like) provide evidence of the potential to investigate soybean yield prediction through its spectral response from multiple platforms of data acquisition, strengthening the possibility of inter-comparison for soybean yield prediction at the R5 phenological stage.

Based on the Sentinel-2 images (global-based model) and Embrapa Soja (Sentinel-2-like model), machine learning regression methods outperformed vegetation indices (Tables 7 and 8), strengthening the potential of spectral bands under machine learning for crop monitoring, as suggested in Tables 4 and 5. However, a better performance from SVR in relation to PLSR was observed. Although PLSR models have been proved to predict yield with high accuracy using hyperspectral data in soybean (Christenson et al., 2016; Crusiol et al., 2021a), the results suggest that, for within-field soybean yield prediction through Sentinel-2 images, the use of SVR can provide higher accuracy, especially when analysing different production areas (fields and farms) into the same dataset.

Although both PLSR and SVR had been successfully applied to yield prediction, the PLSR promotes the linear correlation between predictor variables, i.e., spectral bands, and dependent variable, i.e., soybean yield, while the SVR (the radial basis function) promotes the non-linear correlation between predictors and dependent variable. Chen & Jing (2017) carried out wheat yield estimation through Landsat spectral bands using artificial neural networks and PLSR regression approaches and found that non-linear models provide higher accuracy when integrating data from different fields. Similarly, addressing the soybean yield monitoring using Sentinel-2 images, Dado et al. (2020) demonstrated the effectiveness of non-linear regression models for yield prediction when grouping information from differ-

ent fields. Gómez et al. (2019) carried out potato yield monitoring through Sentinel-2 data and machine learning regression approaches and, although the authors emphasize that in machine learning there is no single algorithm that fits all data, they found that the SVR using a non-linear kernel function (the radial basis function) provided the highest accuracy for yield prediction.

Mapping within-field soybean yield

Towards within-field soybean yield map, field-based models presented the highest accuracy, followed by the farm-based models and, with the lowest accuracy, global-based model. Similar results were obtained in previous studies that addressed within-field yield monitoring based on remote sensing data. Robson et al. (2017) stated that higher correlation between yield and remote sensed data is more likely to be achieved at field scale in comparison to larger monitoring area. Monitoring within-field corn and soybean yield, Kross et al. (2020) reported limitation in the yield prediction and described how the topography from each field, as well as differences in planting, spacing and management practices from each area affected the accuracy of yield prediction within multiple fields. Accordingly, Kayad et al. (2019) suggested that the development and application of corn yield prediction models should be carried out on fields with similar management and crop growth conditions, which might encompass the specific characteristics from each of them.

The global-based model demonstrated lower accuracy for within-field soybean yield prediction, due to the heterogeneous characteristics among fields. However, it has great potential in promoting yield mapping at regional level, considering the average yield from each field or farm. Simultaneously, more information from the crop development conditions within each field, e.g., weather conditions, soil management, invasive plants, disease and insect management, could be useful to strength the prediction ability of the developed models.

The results obtained demonstrate the possibility of developing yield prediction models based on crop spectral response from Sentinel-2 visible, near-infrared and shortwave infrared bands through machine learning approaches, such as SVR. Future research will be developed aiming at exploring multi-source remote sensed data to strengthen the prediction ability of the machine learning regression approaches when analysing yield datasets from multiple fields. Their heterogeneities in planting, spacing and management practices, which might contribute to enhance the design of a soybean within-field yield prediction model with applicability at regional level will be considered.

Conclusions

This research addressed the assessment of strategies for within-field soybean yield monitoring using Sentinel-2 visible, near-infrared and shortwave infrared spectral bands and machine learning regression methods. The results suggest that Sentinel-2 Vis/NIR/SWIR images, associated with partial least squares regression and support vector regression, can be used as a fast and reliable proxy for yield monitoring, contributing to better site-specific management of agronomic practices, economic policies and strategic planning of governmental and corporative decision making over technical issues.

For individual images, single vegetation indices (VIs) performed poorly (low accuracy) in yield prediction compared to the use of the nine Sentinel-2 spectral bands under machine learning regression methods. When pooling all available images across the cropping season, higher accuracies were obtained compared to the use of single images. As has been observed on single images, the use of Vis/NIR/SWIR spectral bands provided higher accuracies than vegetation indices. Besides that, pooling Vis/NIR/SWIR spectral bands and all vegetation indices from all available images did not result in higher accuracy.

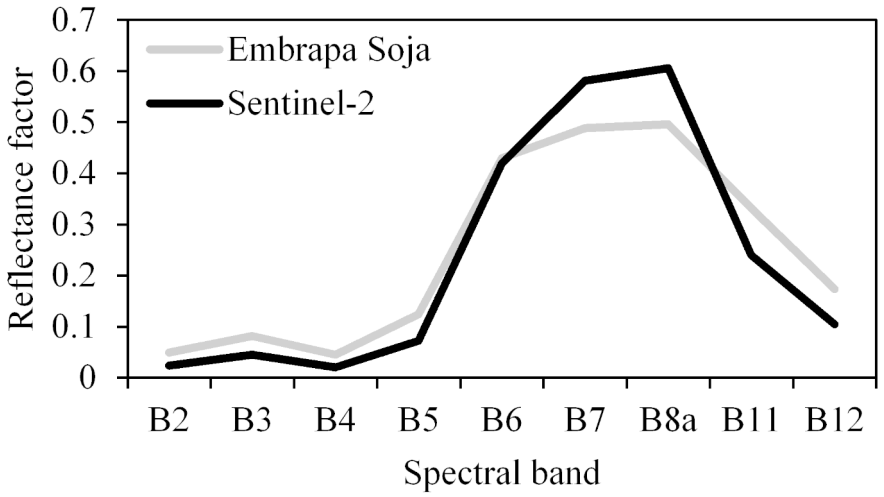
At the R5 phenological stage, with a standardized number of input data (one image containing nine spectral bands), a similar spectral behavior was observed between satellite and ground-based datasets with SVR performing better in relation to PLSR for yield prediction.

For within-field soybean yield mapping based on Sentinel-2 images at the R5 phenological stage, SVR field-based models presented the best results, demonstrating that the accuracy for within-field yield prediction increases from the most complex dataset (global-based model) to models that take into account the complexity from each farm (farm-based model), reaching the highest accuracy using field-based models.

Supplementary Material

Supplementary table Spatial and spectral characteristics of Sentinel-2 spectral bands

		Sentinel-2 A		Sentinel-2 B	
Band number	Spatial resolution (m)	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)
1	60	442.7	21	442.2	21
2	10	492.4	66	492.1	66
3	10	559.8	36	559.0	36
4	10	664.6	31	664.9	31
5	20	704.1	15	703.8	16
6	20	740.5	15	739.1	15
7	20	782.8	20	779.7	20
8	10	832.8	106	832.9	106
8a	20	864.7	21	864.0	22
9	60	945.1	20	943.2	21
10	60	1373.5	31	1376.9	30
11	20	1613.7	91	1610.4	94
12	20	2202.4	175	2185.7	185



Supplementary figure Average soybean spectral response at the R5 phenological stage from the Global-based models (Sentinel-2) and from Embrapa Soja (Sentinel-2-like)

Funding This work was supported by the National Council for Scientific and Technological Development—CNPq; Central Public-Interest Scientific Institution Basal Research Fund [Y2021GH18]; Innovation Project of Chinese Academy of Agricultural Sciences [G202120-5]; and the Talented Young Scientist Program—China Science and Technology Exchange Center [Brazil–19-004].

Availability of data and material (data transparency) Data associated with this research is available with the author L.G.T.C. upon request.

Code availability Not applicable.

Declarations

Conflicts of interest/competing interests The authors declare that they have no conflict of interest.

References

Al-Gaadi, K. A., Hassaballa, A. A., Tola, E., Kayad, A. G., Madugundu, R., Alblewi, B., et al. (2016). Prediction of potato crop yield using precision agriculture techniques. *PloS One*, 11(9). https://doi.org/10.1371/journal.pone.0162219

Ali, A., Martelli, R., Lupia, F., & Barbanti, L. (2019). Assessing multiple years' spatial variability of crop yields using satellite vegetation indices. *Remote Sensing*, 11(20), 2384. https://doi.org/10.3390/rs11202384

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. D. M., & Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, 22(6), 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Ashourloo, D., Aghighi, H., Matkan, A. A., Mobasheri, M. R., & Rad, A. M. (2016). An investigation into machine learning regression techniques for the leaf rust disease detection using hyperspectral measurement. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(9), 4344–4351. https://doi.org/10.1109/JSTARS.2016.2575360

Bolton, D. K., & Friedl, M. A. (2013). Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology*, 173, 74–84. https://doi.org/10.1016/j.agrformet.2013.01.007

Brown, J. C., Kastens, J. H., Coutinho, A. C., de Victoria, D. C., & Bishop, C. R. (2013). Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data. *Remote Sensing of Environment*, 130, 39–50. https://doi.org/10.1016/j.rse.2012.11.009

Carneiro, F. M., Furlani, C. E. A., Zerbato, C., de Menezes, P. C., da Silva Gírio, L. A., & de Oliveira, M. F. (2020). Comparison between vegetation indices for detecting spatial and temporal variabilities in soybean crop using canopy sensors. *Precision Agriculture*, 21, 979–1007. https://doi.org/10.1007/s11119-019-09704-3

Chen, P., & Jing, Q. (2017). A comparison of two adaptive multivariate analysis methods (PLSR and ANN) for winter wheat yield forecasting using Landsat-8 OLI images. *Advances in space research*, 59(4), 987–995. https://doi.org/10.1016/j.asr.2016.11.029

Christenson, B. S., Schapaugh Jr, W. T., An, N., Price, K. P., Prasad, V., & Fritz, A. K. (2016). Predicting soybean relative maturity and seed yield using canopy reflectance. *Crop Science*, 56(2), 625–643. https://doi.org/10.2135/cropsci2015.04.0237

CONAB (National Company of Food Supply) (2021). *Brazilian Crop Assessment–Grain, 2020/2021 Crops, Sixth Inventory Survey, March/2021*. Retrieved March 31, 2021, from https://www.conab.gov.br/info-agro/safras/graos

Crusiol, L. G. T., Carvalho, J. D. F. C., Sibaldelli, R. N. R., Neiverth, W., do Rio, A., Ferreira, L. C., et al. (2017a). NDVI variation according to the time of measurement, sampling size, positioning of sensor and water regime in different soybean cultivars. *Precision Agriculture*, 18(4), 470–490. https://doi.org/10.1007/s11119-016-9465-6

Crusiol, L. G. T., Nanni, M. R., Furlanetto, R. H., Sibaldelli, R. N. R., Cezar, E., Sun, L., et al. (2021a). Yield Prediction in Soybean Crop Grown under Different Levels of Water Availability Using Reflectance Spectroscopy and Partial Least Squares Regression. *Remote Sensing*, 13(5), 977. https://doi.org/10.3390/rs13050977

Crusiol, L. G. T., Nanni, M. R., Furlanetto, R. H., Sibaldelli, R. N. R., Cezar, E., Sun, L., et al. (2021b). Classification of Soybean Genotypes Assessed Under Different Water Availability and at Different Phenological Stages Using Leaf-Based Hyperspectral Reflectance. *Remote Sensing*, 13(2), 172. https://doi.org/10.3390/rs13020172

Crusiol, L. G. T., Neto, O. C. P., Nanni, M. R., da Silva Gualberto, A. A., Furlanetto, R. H., & da Silva Junior, C. A. (2017b). Mapeamento de áreas agrícolas na safra de verão a partir de imagens Landsat frente aos dados oficiais (Mapping of agricultural areas in the summer crop season using Landsat images against official data). *Revista Agro@mbiente On-line, 10*(4), 287–298. https://doi.org/10.18227/1982-8470ragro.v10i4.3098

Dado, W. T., Deines, J. M., Patel, R., Liang, S. Z., & Lobell, D. B. (2020). High-Resolution Soybean Yield Mapping Across the US Midwest Using Subfield Harvester Data. *Remote Sensing*, 12(21), 3471. https://doi.org/10.3390/rs12213471

Di Gennaro, S. F., Dainelli, R., Palliotti, A., Toscano, P., & Matese, A. (2019). Sentinel-2 validation for spatial variability assessment in overhead trellis system viticulture versus UAV and agronomic data. *Remote Sensing*, 11(21), 2573. https://doi.org/10.3390/rs11212573

Eberhardt, I. D. R., Schultz, B., Rizzi, R., Sanches, I. D. A., Formaggio, A. R., Atzberger, C., et al. (2016). Cloud cover assessment for operational crop monitoring systems in tropical areas. *Remote Sensing*, 8(3), 219. https://doi.org/10.3390/rs8030219

Embrapa Soja. (2013). *Tecnologias de Produção de Soja—Região Central do Brasil 2014 (Technologies for Soybean Production—Central Region of Brazil 2014); Embrapa Soja*. Brazil: Londrina

ESA—The European Space Agency. *Sentinel-2 User Guide*. Retrieved March 31 (2021). from https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi

Esquerdo, J. C. D. M., Zullo Júnior, J., & Antunes, J. F. G. (2011). Use of NDVI/AVHRR time-series profiles for soybean crop monitoring in Brazil. *International Journal of Remote Sensing*, 32(13), 3711–3727. https://doi.org/10.1080/01431161003764112

FAO (Food and Agriculture Organization of the United Nations) (2018). *The future of food and agriculture—Alternative pathways to 2050*. Summary version. Rome. 60 pp. Licence: CC BY-NC-SA 3.0 IGO. Retrieved March 31, 2021 from http://www.fao.org/3/I8429EN/i8429en.pdf. Accessed on 31 March 2021

Fehr, W. R., & Caviness, C. E. (1977). Stages of Soybean Development; *Special Report 80*; Iowa State University of Science and Technology: Ames, IA, USA

Gao, F., Anderson, M., Daughtry, C., & Johnson, D. (2018). Assessing the variability of corn and soybean yields in central Iowa using high spatiotemporal resolution multi-satellite imagery. *Remote Sensing*, 10(9), 1489. https://doi.org/10.3390/rs10091489

Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote sensing of Environment*, 58(3), 289–298. https://doi.org/10.1016/S0034-4257(96)00072-7

Gitelson, A., & Merzlyak, M. N. (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. https://doi.org/10.1016/S0176-1617(11)81633-0

Gómez, D., Salvador, P., Sanz, J., & Casanova, J. L. (2019). Potato yield prediction using machine learning techniques and sentinel 2 data. *Remote Sensing*, 11(15), 1745. https://doi.org/10.3390/rs11151745

Guan, K., Wu, J., Kimball, J. S., Anderson, M. C., Frolking, S., Li, B., et al. (2017). The shared and unique values of optical, fluorescence, thermal and microwave satellite data for estimating large-scale crop yields. *Remote Sensing of Environment*, 199, 333–349. https://doi.org/10.1016/j.rse.2017.06.043

Gusso, A., & Ducati, J. R. (2012). Algorithm for soybean classification using medium resolution satellite images. *Remote Sensing*, 4(10), 3127–3142. https://doi.org/10.3390/rs4103127

Hardisky, M. A., Klemas, V., & Smart, M. (1983). The influence of soil salinity, growth form, and leaf moisture on the spectral radiance of spartina alterniflora canopies. *Photogrammetric Engineering and Remote Sensing*, 49, 77–83

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1–2), 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2

Hunt, M. L., Blackburn, G. A., Carrasco, L., Redhead, J. W., & Rowland, C. S. (2019). High resolution wheat yield mapping using Sentinel-2. *Remote Sensing of Environment*, 233, 111410. https://doi.org/10.1016/j.rse.2019.111410

Jiang, Z., Huete, A. R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, 112(10), 3833–3845. https://doi.org/10.1016/j.rse.2008.06.006, https://doi.org/10.1016/j.rse.2013.10.027

Kamir, E., Waldner, F., & Hochman, Z. (2020). Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160, 124–135. https://doi.org/10.1016/j.isprsjprs.2019.11.008

Kayad, A., Sozzi, M., Gatto, S., Marinello, F., & Pirotti, F. (2019). Monitoring within-field variability of corn yield using Sentinel-2 and machine learning techniques. *Remote Sensing*, 11(23), 2873. https://doi.org/10.3390/rs11232873

Kross, A., Znoj, E., Callegari, D., Kaur, G., Sunohara, M., Lapen, D. R., et al. (2020). Using Artificial Neural Networks and Remotely Sensed Data to Evaluate the Relative Importance of Variables for Prediction of Within-Field Corn and Soybean Yields. *Remote Sensing*, 12(14), 2230. https://doi.org/10.3390/rs12142230

Lamquin, N., Woolliams, E., Bruniquel, V., Gascon, F., Gorroño, J., Govaerts, Y., et al. (2019). An inter-comparison exercise of Sentinel-2 radiometric validations assessed by independent expert groups. *Remote Sensing of Environment*, 233, 111369

Leon, C. T., Shaw, D. R., Cox, M. S., Abshire, M. J., Ward, B., Wardlaw, M. C., et al. (2003). Utility of remote sensing in predicting crop and soil characteristics. *Precision Agriculture*, 4(4), 359–384. https://doi.org/10.1023/A:1026387830942

Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., & Fritschi, F. B. (2020). Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sensing of Environment*, 237, 111599. https://doi.org/10.1016/j.rse.2019.111599

Mercante, E., Lamparelli, R. A., Uribe-opazo, M. A., & Rocha, J. V. (2010). Modelos de regressão lineares para estimativa de produtividade da soja no oeste do Paraná, utilizando dados espectrais (Linear regression models to soybean yield estimate in the west region of the State of Paraná, Brazil, using spectral data). *Engenharia Agrícola*, 30(3), 504–517. https://doi.org/10.1590/S0100-69162010000300014

Munyati, C., Balzter, H., & Economon, E. (2020). Correlating Sentinel-2 MSI-derived vegetation indices with in-situ reflectance and tissue macronutrients in savannah grass. *International Journal of Remote Sensing*, 41(10), 3820–3844. https://doi.org/10.1080/01431161.2019.1708505

QGIS (2022). QGIS Geographic Information System. QGIS Association. http://www.qgis.org

Rahman, M. M., Robson, A., & Bristow, M. (2018). Exploring the potential of high resolution worldview-3 Imagery for estimating yield of mango. *Remote Sensing*, 10(12), 1866. https://doi.org/10.3390/rs10121866

Robson, A., Rahman, M. M., & Muir, J. (2017). Using worldview satellite imagery to map yield in avocado (Persea americana): A case study in Bundaberg, Australia. *Remote Sensing*, 9(12), 1223. https://doi.org/10.3390/rs9121223

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

Sakamoto, T. (2020). Incorporating environmental variables into a MODIS-based crop yield estimation method for United States corn and soybeans through the use of a random forest regression algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160, 208–228. https://doi.org/10.1016/j.isprsjprs.2019.12.012

SEAB (Secretaria de Estado da Agricultura e do Abastecimento do Paraná—Departamento de Economia Rural) (2021). *Paraná State Crop Assessment, 2020/2021*. Retrieved March 31, 2021, from https://www.conab.gov.br/info-agro/safras/graos

Segarra, J., González-Torralba, J., Aranjuelo, Í., Araus, J. L., & Kefauver, S. C. (2020). Estimating Wheat Grain Yield Using Sentinel-2 Imagery and Exploring Topographic Features and Rainfall Effects on Wheat Performance in Navarre, Spain. *Remote Sensing*, 12(14), 2278. https://doi.org/10.3390/rs12142278

Shoko, C., & Mutanga, O. (2017). Examining the strength of the newly-launched Sentinel 2 MSI sensor in detecting and discriminating subtle differences between C3 and C4 grass species. *ISPRS Journal of Photogrammetry and Remote Sensing*, 129, 32–40. https://doi.org/10.1016/j.isprsjprs.2017.04.016

Silva Junior, C. A., da, Nanni, M. R., Teodoro, P. E., & Silva, G. F. C. (2017). Vegetation indices for discrimination of soybean areas: A New Approach. *Agronomy Journal*, 109(4), 1331–1343. https://doi.org/10.2134/agronj2017.01.0003

Silva, E. E., da, Baio, F. H. R., Teodoro, L. P. R., da Silva Junior, C. A., Borges, R. S., & Teodoro, P. E. (2020). UAV-multispectral and vegetation indices in soybean grain yield prediction based on in situ observation. *Remote Sensing Applications: Society and Environment*, 18, 100318. https://doi.org/10.1016/j.rsase.2020.100318

Souza, A. M. D., Breitkreitz, M. C., Filgueiras, P. R., Rohwedder, J. J. R., & Poppi, R. J. (2013). Experimento didático de quimiometria para calibração multivariada na determinação de paracetamol em comprimidos comerciais utilizando espectroscopia no infravermelho próximo: um tutorial, parte II (Teaching experiment of chemometrics for multivariate calibration in determination of paracetamol in commercial tablets using near-infrared spectroscopy: a tutorial, part II). *Química Nova*, 36(7), 1057–1065. https://doi.org/10.1590/S0100-40422013000700022

Suarez, L. A., Robson, A., McPhee, J., O'Halloran, J., & van Sprang, C. (2020). Accuracy of carrot yield forecasting using proximal hyperspectral and satellite multispectral data. *Precision Agriculture*, 21(6), 1304–1326. https://doi.org/10.1007/s11119-020-09722-6

Sugawara, L. M., Rudorff, B. F. T., & Adami, M. (2008). Viabilidade de uso de imagens do Landsat em mapeamento de área cultivada com soja no Estado do Paraná (Feasibility of the use of Landsat imagery to map soybean crop areas in Paraná, Brazil). *Pesquisa Agropecuária Brasileira*, 43(12), 1777–1783

Toming, K., Kutser, T., Laas, A., Sepp, M., Paavel, B., & Nõges, T. (2016). First experiences in mapping lake water quality parameters with Sentinel-2 MSI imagery. *Remote Sensing*, 8(8), 640. https://doi.org/10.3390/rs8080640

USDA (United States Department of Agriculture) (2021). *World Agricultural Production. Circular Series WAP 3–21, March 2021*. Retrieved March 31, 2021 from https://apps.fas.usda.gov/psdonline/circulars/production.pdf

Vega, A., Córdoba, M., Castro-Franco, M., & Balzarini, M. (2019). Protocol for automating error removal from yield maps. *Precision Agriculture*, 20(5), 1030–1044. https://doi.org/10.1007/s11119-018-09632-8

Wang, F. M., Huang, J. F., Tang, Y. L., & Wang, X. Z. (2007). New vegetation index and its application in estimating leaf area index of rice. *Rice Science*, 14(3), 195–203. https://doi.org/10.1016/S1672-6308(07)60027-4

Wrege, M. S., Steinmetz, S., Reiser, C. Jr., & de Almeida, I. R. (2011). *Atlas Climático da Região Sul do Brasil: Estados do Paraná, Santa Catarina e Rio Grande do Sul (Climate atlas from the south region of Brazil: States of Paraná, Santa Catarina and Rio Grande do Sul); Embrapa Clima Temperado*. Colombo, Brazil: Pelotas, Brazil; Embrapa Florestas

Yendrek, C. R., Tomaz, T., Montes, C. M., Cao, Y., Morse, A. M., Brown, P. J., et al. (2017). High-throughput phenotyping of maize leaf physiological and biochemical traits using hyperspectral reflectance. *Plant Physiology*, 173(1), 614–626. https://doi.org/10.1104/pp.16.01447

Zhai, Z., Martínez, J. F., Beltran, V., & Martínez, N. L. (2020). Decision support systems for agriculture 4.0: Survey and challenges. *Computers and Electronics in Agriculture*, 170, 105256. https://doi.org/10.1016/j.compag.2020.105256

Zhang, X., Zhao, J., Yang, G., Liu, J., Cao, J., Li, C., et al. (2019). Establishment of plot-yield prediction models in soybean breeding programs using UAV-based hyperspectral remote sensing. *Remote Sensing*, 11(23), 2752. https://doi.org/10.3390/rs11232752

Authors and Affiliations

L. G.T. Crusiol[1,2] · Liang Sun[1] · R. N.R. Sibaldelli[3] · V. Felipe Junior[4] · W. X. Furlaneti[4] · R. Chen[1] · Z. Sun[1] · D. Wuyun[1] · Z. Chen[5] · M. R. Nanni[2] · R. H. Furlanetto[2] · E. Cezar[2] · A. L. Nepomuceno[3] · J. R.B. Farias[3]

✉ Liang Sun
sunliang@caas.cn

L. G.T. Crusiol
luiscrusiol@gmail.com

R. N.R. Sibaldelli
rubson.sibaldelli@embrapa.br

V. Felipe Junior
vanderlei.junior@integrada.coop.br

W. X. Furlaneti
wellington.furlaneti@integrada.coop.br

R. Chen
chenrq@mails.ccnu.edu.cn

Z. Sun
sunzhengcaas@gmail.com

D. Wuyun
82101181154@caas.cn

Z. Chen
zhongxin.chen@fao.org

M. R. Nanni
mrnanni@uem.br

R. H. Furlanetto
renatohfurlanetto@hotmail.com

E. Cezar
eccarpejani@gmail.com

A. L. Nepomuceno
alexandre.nepomuceno@embrapa.br

J. R.B. Farias
joserenato.farias@embrapa.br

[1] Key Laboratory of Agricultural Remote Sensing, Ministry of Agriculture/CAAS-CIAT Joint Laboratory in Advanced Technologies for Sustainable Agriculture—Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, 100081 Beijing, China

[2] Department of Agronomy, State University of Maringá, 87020-900 Maringá, PR, Brazil

[3] Embrapa Soja (National Soybean Research Center—Brazilian Agricultural Research Corporation), 86001-970 Londrina, PR, Brazil

[4] Integrada Cooperativa Agroindustrial, 86010-480 Londrina, PR, Brazil

[5] Digitalization and Informatics Division, Food and Agricultural Organization of the United Nations, Terme Caracalla, 00153 Rome, Italy