AgriEngineering

Article

Coffee-Yield Estimation Using High-Resolution Time-Series Satellite Images and Machine Learning

Maurício Martello *, José Paulo Molin, Marcelo Chan Fu Wei, Ricardo Canal Filho and João Vitor Moreira Nicoletti

Laboratory of Precision Agriculture (LAP), Department of Biosystems Engineering, "Luiz de Queiroz" College of Agriculture (ESALQ), University of Sao Paulo (USP), Piracicaba 13418-900, Brazil
* Correspondence: mauriciomartello@usp.br

Abstract: Coffee has high relevance in the Brazilian agricultural scenario, as Brazil is the largest producer and exporter of coffee in the world. Strategies to advance the production of coffee grains involve better understanding its spatial variability along fields. The objectives of this study were to adjust yield-prediction models based on a time series of satellite images and high-density yield data, and to indicate the best phenological stage of coffee crop to obtain satellite images for this purpose. The study was conducted during three seasons (2019, 2020 and 2021) in a commercial area (10.24 ha), located in the state of Minas Gerais, Brazil. Data were obtained using a harvester equipped with a yield monitor that measures the volume of coffee harvested with 3.0 m of spatial resolution. Satellite images from the PlanetScope (PS) platform were used. Random forest (RF) regression and multiple linear regression (MLR) models were fitted to different datasets composed of coffee yield and time series of satellite-image data ((1) Spectral bands—red, green, blue and near-infrared; (2) Normalized difference vegetation index (NDVI); or (3) Green normalized difference vegetation index (GNDVI)). Whether using RF or MLR, the spectral bands, NDVI and GNDVI reproduced the spatial variability of yield maps one year before harvest. This information can be of critical importance for management decisions across the season. For yield quantification, the RF model using spectral bands showed the best results, reaching R^2 of 0.93 for the validation set, and the lowest errors of prediction. The most appropriate phenological stage for satellite-image data acquisition was the dormancy phase, observed during the dry season months of July and August. These findings can help to monitor the spatial and temporal variability of the fields and guide management practices based on the premises of precision agriculture.

Keywords: remote sensing; precision agriculture; NDVI; GNDVI

check for updates

Citation: Martello, M.; Molin, J.P.; Wei, M.C.F.; Canal Filho, R.; Nicoletti, J.V.M. Coffee-Yield Estimation Using High-Resolution Time-Series Satellite Images and Machine Learning. *AgriEngineering* **2022**, *4*, 888–902. https://doi.org/10.3390/agriengineering4040057

Academic Editors: Luis A. Ruiz and Francesco Marinello

Received: 13 August 2022
Accepted: 30 September 2022
Published: 5 October 2022

1. Introduction

A coffee crop interacts with several factors, such as soil, climate and the plant itself, presenting particularities that result in the high spatial variability of yield [1,2]. In this sense, delineating high- and low-yielding zones can guide farmers to the adoption of management strategies across the season [3–5]. However, knowledge about yield before harvest is still a challenge [6].

Remote sensing (RS) is a potential source of data for monitoring agricultural fields. At the orbital level, RS [7] stands out for its high coverage rate, allowing the monitoring of large areas combined with a temporal database. Recent research has demonstrated that spectral variations captured by orbital imaging can be related to soil and crop characteristics, identifying patterns of interest for agriculture [1,8,9]. These data are a potential support for the diagnosis of agronomic parameters and decision making in agricultural production steps [10–13]. However, some of the main limitations related to orbital images are the lack of field data that, in synergy with remote sensing techniques, will enable the calibration of predictive models [14]. Furthermore, researchers showed that agronomic prediction models may have spatiotemporal constraints for application in different fields and seasons [15–19]. This reinforces the necessity to establish simple, fast and cost-effective methods to calibrate prediction models in the different fields of agriculture.

The assessment of crop spatial variability is challenging due to the limited solutions available, mainly for yield mapping in coffee crops. Yield can be considered the most important data layer to start the investigation of spatial variability within the field [20], to delimit management zones and enhance site-specific applications [16,21,22], leveraging precision-agriculture (PA) management strategies.

Currently, the coffee-crop yield, by hand or machine harvesting, may be estimated from in-field samples [23–25] or exclusively with mechanical harvesting, using a yield monitor [26,27]. Sampling coffee fruits in-field is an onerous and destructive process, besides its difficult implementation for large areas [28]. A commercially available yield monitor measures the volume of harvested fruits using a conveyor belt with cells of known volume [26,27,29]. This method is associated with a high investment cost and its use is exclusive to a single brand of coffee harvester. In addition, yield maps require calibration, the accessibility of data-processing tools and knowledge to manage data from multiple harvesters.

Studies have reported the strength of the linear correlation between vegetation indices (VIs) using orbital images and coffee-yield prediction. Bernardes et al. [30] evaluated possible correlations between coffee yield and MODIS-derived vegetation indices, but the authors obtained the yield data interviewing producers, with no direct field acquisition. Nogueira et al. [31] reported the use of vegetation indices obtained with images from the Landsat-8 satellite's OLI (operational land imager) sensor to estimate yield. Evaluating two seasons considered low- and high-yielding, they found that NDVI had the strongest yield correlation during the dormancy and flowering stages. Thao et al. [32], in a study that attempted to assess a coffee-yield forecasting at Dak Lak province in Vietnam using vegetation indexes (NDVI, LAI and FAPAR) derived from SPOT-VEGETATION and PROBA-V satellites obtained satisfactory accuracy (Adj. R^2 = 64 to 69%) in estimating yield in coffee fields by means of multiple linear regression models using data from the first semester of the years 2000–2019.

Recently, Silva et al. [33] evaluated the correlation of different phenological stages' spectral responses of coffee in a center pivot and concluded that the method was suitable for predicting coffee yield. However, the strategy applied was of manual sampling points, then extrapolating them to the whole field. This strategy of low-density yield samples used as ground-truth data is widely observed for coffee crops due to the difficult implementation of high-density yield-data acquisition. However, the low density can compromise the ability of the model to trustworthily represent yield spatial variability.

Aware of the lack of studies based on remote sensing with yield-monitor data for coffee crops [1], it is reasonable to test a solution proposed for sugarcane crops [34] with coffee. The method to estimate sugarcane yield is based on satellite images and yield monitor data at high density, using machine-learning (ML) techniques and random forest (RF) regression, which can cope with both linear and non-linear relation and which have higher prediction accuracy compared to standard statistics (i.e., multiple linear regression that performs better on linear relations), and allow us to identify the best datasets to be used in yield estimation [35–37].

Therefore, the objective of this work was: (a) to adjust yield-prediction models based on a time series of satellite-image data (spectral bands and VIs) and high-density data from yield monitors using random forest and multiple linear regression algorithms, and (b) indicate the appropriate phenological stage of the coffee crop to obtain satellite-image data for this purpose.

2. Materials and Methods

The study was conducted in a commercial area located in the municipality of Patos de Minas, Minas Gerais state, Brazil, with central geographic coordinates at 18°32′28.55″ S latitude and 46°3′51.17″ W longitude (coordinate reference system WGS 84) and an altitude of 1025 m (Figure 1). The climate of the study area is classified as Aw, tropical with dry winter and rainy summer, according to the Köppen climate classification [38]. The monthly normal temperatures (1991–2020) range from 18.9 °C in the coldest month (June) to 23.4 °C in the warmest month (October) with average annual temperature of 21.6 °C [38]. The area of interest had 10.24 ha cultivated with the species *Coffea arabica* L. (IAC Catuaí 144 variety) planted in 2006 and had its first harvest in 2009, under a drip irrigation system, with yields ranging from 0.98 to 2.61 Mg ha^{-1} during the evaluated period.

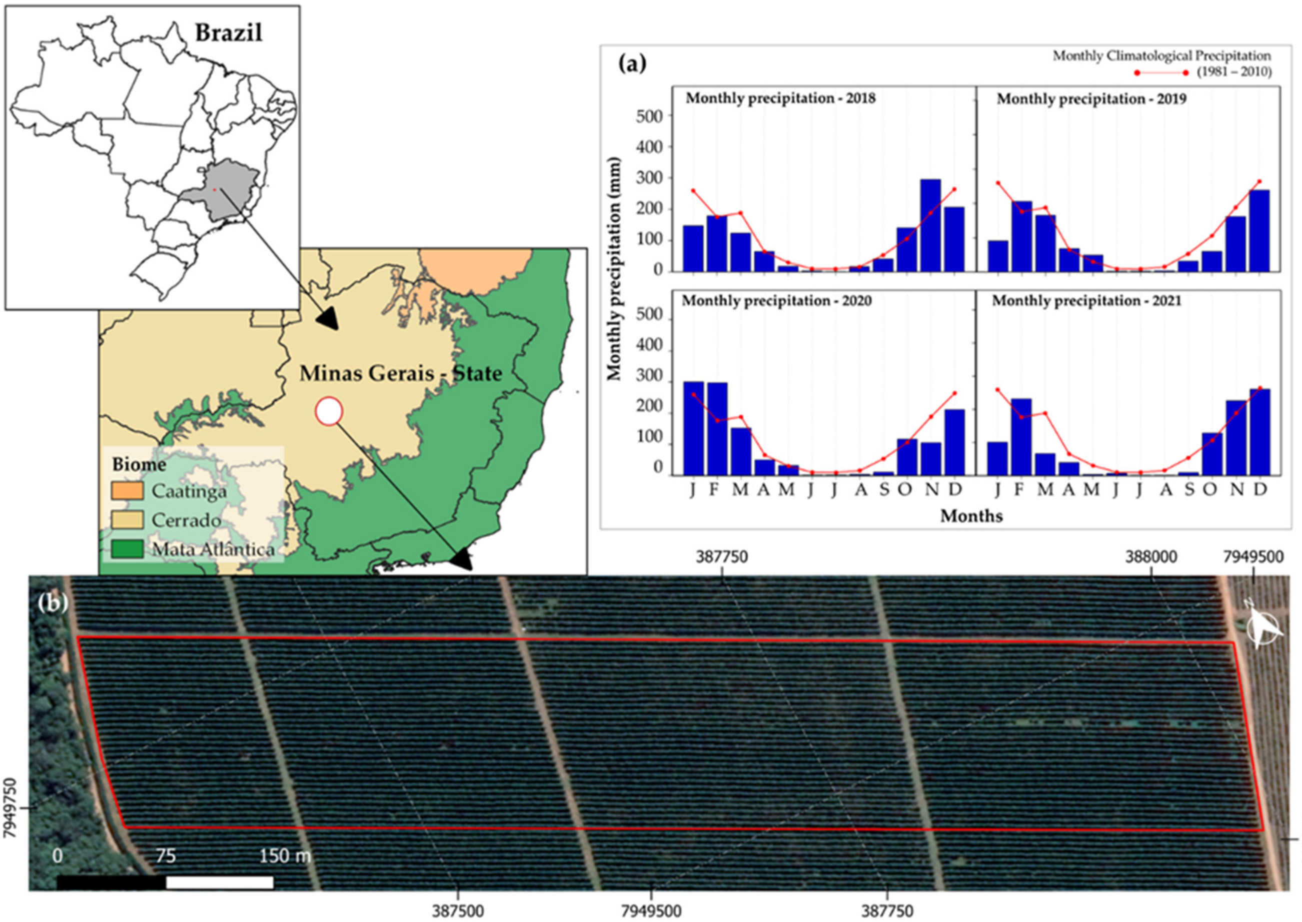

Figure 1. Location of coffee-field study site in Brazil. (**a**) Historical data of monthly precipitation for the years 2018, 2019, 2020 and 2021 and the line with the monthly climatological precipitation of 30 years (1981–2010) for the study area, [39]. (**b**) Study area; the red line represents the area boundary.

2.1. Field Data

Data were obtained using a K3 Millennium harvester (Jacto, Pompeia, Brazil), equipped with a yield monitor that measures the volume of coffee harvested with a resolution of approximately 3.0 m for the total area. Full information about the methodology and its proper validation and reliability assessment are available in Martello et al. [27]. The data were collected during the 2019, 2020 and 2021 harvests, then converted to weight of processed coffee through a conversion factor [27]. Headland-maneuver and bordering-area data close to roads that divide and cross the area were first removed. Discrepant data inside the field were filtered using the MapFilter 2.0 software using global filtering with a threshold of 100%, based on the methodology proposed by Maldaner and Molin [40]. Yield data were interpolated using the Vesper software 1.6 [41] using the ordinary kriging method with a spatial resolution of 3.0 m × 3.0 m, and the grid obtained from the central coordinates of the satellite-image pixels.

2.2. Satellite Data

Satellite images from PlanetScope (PS) were used [42]. PS images had 3.0 m of spatial resolution with cloud-free coverage, and each image included four spectral bands: blue (455–515 nm), green (500–590 nm), red (590–670 nm) and near-infrared (780–860 nm). Considering the condition of zero cloud coverage, it 33 images were obtained between 2018 and 2021, always at the end of the month and trying to keep an interval of 30 days between images from the sensor Dove Classic—PS2, product Orto Scene—Analytic—Level 3B; these images underwent a series of processes by the vendor, including sensor and radiometric correction, atmospheric correction and conversion to top-of-atmosphere reflectance (TOA) and geometric correction (Table 1).

Table 1. Dates of the PlanetScope orbital images obtained.

Harvest 1st (2018/19)	Harvest 2nd (2019/20)	Harvest 3rd (2020/21)
30 July 18	28 July 19	27 July 20
31 August 18	31 August 19	31 August 20
29 September 18	22 September 19	28 September 20
27 October 18	14 October 19	26 October 20
21 December 18	31 December 19	18 December 20
30 January 19	14 January 20	30 January 21
25 February 19	21 February 20	1 February 21
15 March 19	19 March 20	21 March 21
27 April 19	25 April 20	28 April 21
28 May 19	28 May 20	25 May 21
30 June 19	28 June 20	28 June 21

No images were found without cloud cover in the month of November for the three years.

2.3. Statistical Analysis

The dataset was composed of high-density samples of coffee yield, spectral bands (red, green, blue and NIR); the normalized difference vegetation index (NDVI) was proposed by Rouse et al. [43] and the calculation was performed using the normalized difference between the spectral regions of red and near infrared, showing correlation with the green biomass of the plants [44]. Additionally, the green normalized difference vegetation index (GNDVI), which was proposed by Gitelson et al. [45], was used, using the green band instead of the red band which increases the sensitivity of the index in identifying the concentration of chlorophyll when compared to the NDVI [44].

Before fitting yield models, the Pearson correlation coefficient (r) was calculated among variables (spectral bands × yield, NDVI × yield and GNDVI × yield) aiming to find those that present the highest linear correlation values with yield and indicate the most suitable periods to obtain satellite-imagery data, while being aware of the cloud-cover and temporal resolution limitations. To find the most suitable period the use of Vis was chosen, since they can be considered a dimensionality-reduction method which might facilitate the interpretation of the data correlation.

RF regression and MLR were fitted to the different datasets composed of the temporal series (spectral bands and VIs) and the selected periods based on the indication of the most suitable period (VIs).

According to Breiman [46], RF is an algorithm composed of several decision trees, where each tree depends on the values of a random vector sampled independently of the input vector with an identical distribution for all trees within the forest. In this study, RF regression was implemented in RStudio (R Core Team, 2018) using the "randomForest" package [47]. The coffee-yield predicted value is the mean fitted response from all the individual trees that resulted from each bootstrapped sample.

MLR is a regression method that aims at one target variable related to multiple features, where the target can be estimated using Equation (1) [48].

$$Y = X\beta + e \tag{1}$$

where Y is a (n × 1) target vector, X is a (n × p) features matrix (predictor variables), β is a p × 1 vector of unknown coefficients and e is a n × 1 random vector of errors.

Yield-predictive models were compared, considering the coefficient of determination (R^2), root mean squared error (RMSE—Equation (2)) and mean absolute error (MAE—Equation (3)). These parameters were calculated for training (2/3), test (1/3) and the entire dataset (3/3). Yield maps were generated using the geographic information system (GIS) Quantum GIS—QGIS [49].

$$RMSE = \left\{ n^{-1} \left[\sum (y_i - \hat{y})^2 + \ldots + (y_n - \hat{y})^2 \right] \right\}^{0.5} \quad (2)$$

where $RMSE$ = root mean squared error, n = number of samples, y = observed variable response and $\hat{y}$ = predicted variable response.

$$MAE = \left\{ n^{-1} \left[\sum (|y_i - \hat{y}|) + \ldots + (|y_n - \hat{y}|) \right] \right\} \quad (3)$$

where MAE = mean absolute error, n = number of samples, y = observed variable response and $\hat{y}$ = predicted variable response.

A flowchart corresponding to the coffee-yield prediction and mapping procedure is shown in Figure 2. It presents the process through the stages of data collection using satellite imagery (including pre-processing and data selection), georeferenced coffee-yield sampling, data merging (satellite imagery and coffee-yield sampling data), data splitting (train and test data) and RF regression e MLR application.

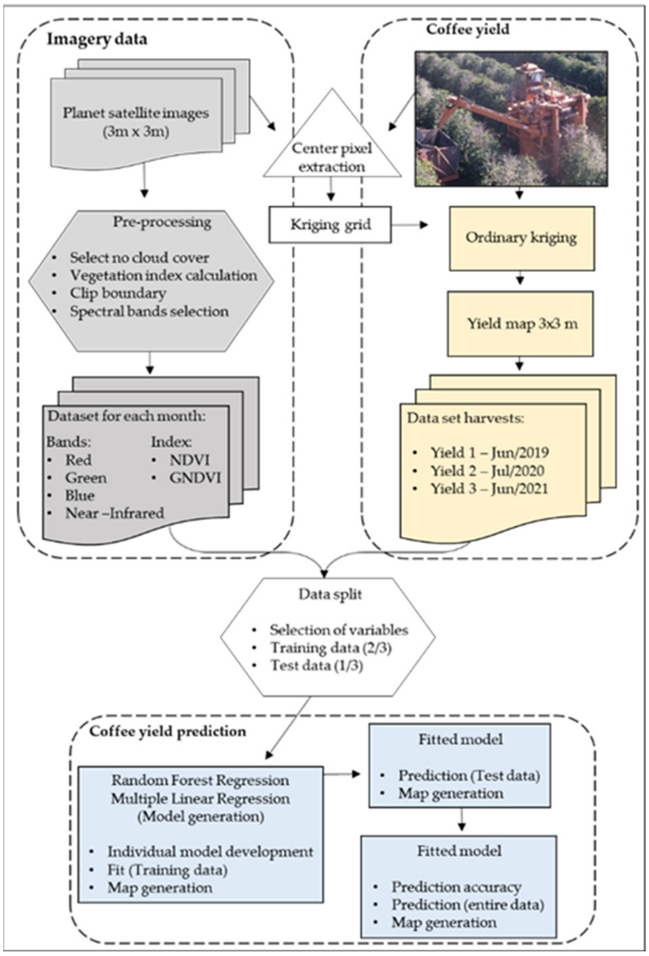

Figure 2. Coffee-yield prediction and mapping flowchart.

3. Results and Discussion

Figure 3 shows the Pearson correlation values among spectral bands, VIs and yield from 2019, 2020 and 2021 harvests. The highest correlation values were found one year before harvest regardless of the season. July and August were the months that presented the highest correlation values with yield, a fact related to the plant phenology and already reported by some authors, since the vegetative vigor of the plant can be inferred from this stage [30,31,33].

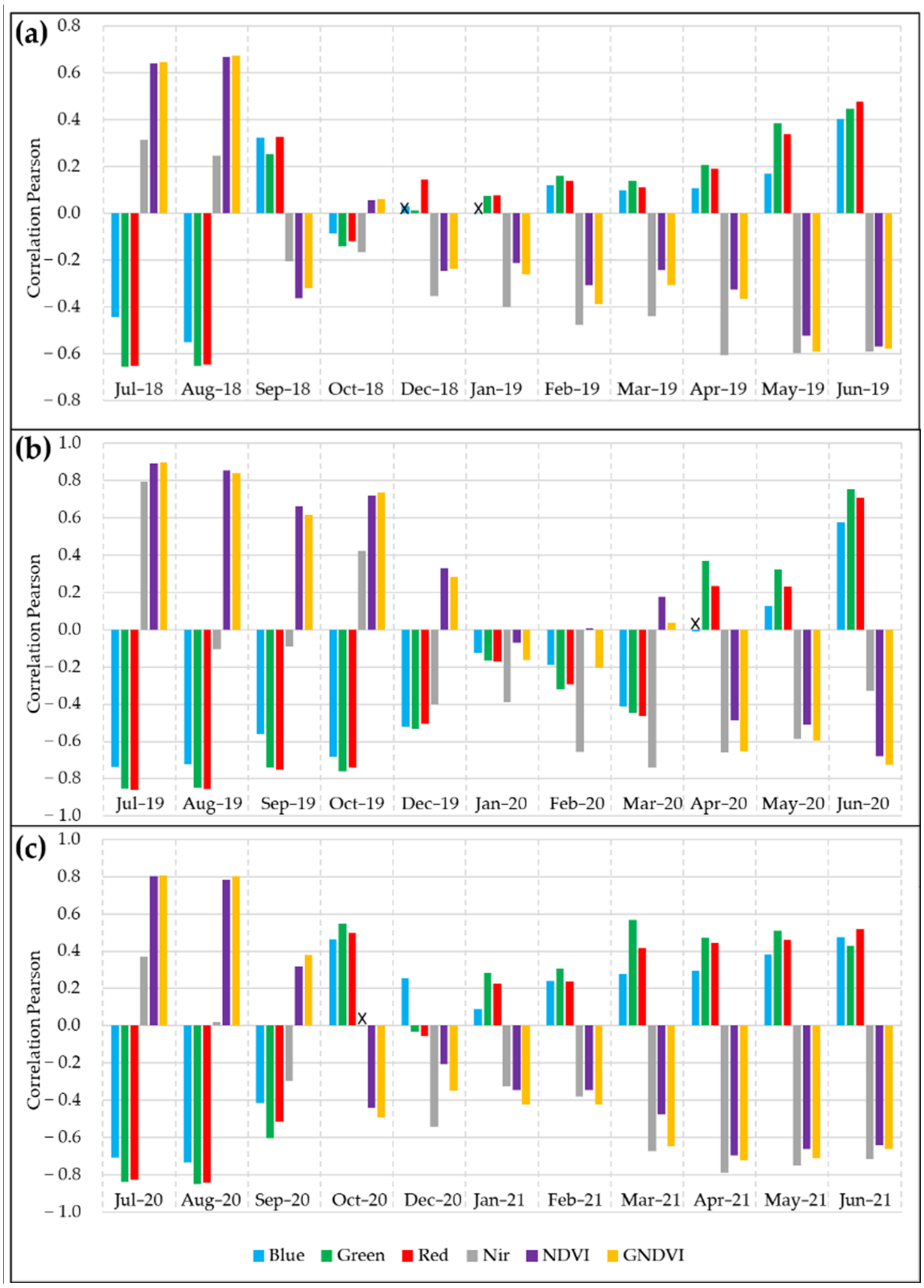

Figure 3. Pearson correlation between spectral bands/vegetative indices and coffee yield at a significance level of 5%. Cross-marks (X) mean that the variables were not significant. Nir: Near infrared, NDVI: Normalized Difference Vegetation Index, GNDVI: Green Normalized Difference Vegetation Index. (**a**) 2019 harvest; (**b**) 2020 harvest; (**c**) 2021 harvest.

For the 2019 harvest (performed in June 2019), the highest values of r were 0.64 and 0.65 for NDVI and GNDVI on July and 0.67 for both NDVI and GNDVI for August. In the 2020 harvest (performed in July 2020), the highest r values were 0.89 and 0.85 for NDVI and GNDVI, respectively, on July and 0.85 for NDVI and 0.83 for GNDVI in August. Note that a similar behavior was found for 2021's harvest (June 2021), presenting r values of 0.8 and 0.78 for NDVI and GNDVI on July and 0.8 for both NDVI and GNDVI in August.

July and August were also the most correlated months, indicated by Silva et al. (2021) [33] and could be used to predict coffee yield. These months are related to the phenological stage (dormancy bud phase) of the plant on which potential production of the next year is already established due to flowering induction occurring in the previous months [50]. The coffee plant enters a mandatory dormancy stage during the period of water deficit (months of dry winter in Brazil) until the rainfall breaks flower-bud dormancy [51,52].

The coffee crop shows intense vegetative growth for one year, allowing it to produce grains more intensively (reproductive stage) in the coming year [53]. During the dormancy phase, plagiotropic branches (productive branches) go into senescence. Due to the vegetative growth stage, orthotropic branches (vegetative stage) receive more nutrients than the plagiotropic branches until the next reproductive stage so they can form new branches and leaves [51,54,55]. The alternation of r values (negative and positive) in the temporal sequence can be explained based on the phenological process of the crop, where positive values can be found right after the prior harvest season and negative values during the development of the crop. These results indicate that the dormancy phase (July and August) based on VIs can be an indicative of potential yield in qualitative terms for the next harvest.

Based on the results from NDVI and/or GNDVI, it can be inferred that yield modelling could be conducted using only two months (July and August) for linear models, in case of limited data from satellite imagery, which brings benefits since these months constitute the dry season, and are thus not greatly limited by the interference of cloud cover.

From the results of the linear correlation, yield-predictive models were fitted according to the different types of datasets: (a) temporal series of spectral bands, NDVI and GNDVI and (b) data from July and August for NDVI and GNDVI. It was selected the use of July and August because they presented the highest Pearson correlation with yield.

Table 2 shows RMSE, R^2 and MAE results for training, test and the entire dataset, applying RF and MLR to predict coffee yield based on different types of variables and months within seasons. Note that the highest R^2 results are found in the models based on the RF regression regardless of the dataset, as also found in Wei et al. [56] and Canata et al. [34]. These results highlight that even if there are some linear correlations among yield and satellite-imagery data (spectral bands or vegetation index), the RF in this context is likely to be used instead of MLR regression.

Comparing the results relying on the temporal series (11 months) and only on the two best months (July and August), it can be noted that the fewer variables used, the lower the accuracy regardless of the dataset and regression model.

Figures 4–6 represent the yield maps from the yield monitor and different yield-predictive models. From Figures 4A, 5A and 6A, the biennial bearing effect of coffee yield can be visualized, an expected phenomenon and one widely reported in the literature [30,33] in which harvest 1 and 3 could be considered as low-yielding seasons and harvest 2 as a high-yielding season. However, the effect is usually reported in studies performed with sampling at low spatial resolution, which is different from this work in that it shows results based on the entire harvest with high-density data. Thus, from these results not only the variability between years can be inferred, but also the variability within the field, which can provide data for decisions makers to improve crop management [27,29,57].

Table 2. Mean squared error (RMSE), coefficient of determination (R^2), mean absolute error (MAE) results for training, test and the entire dataset, applying random forest and multiple linear regression to predict coffee yield based on different types of variables and months within harvests.

Variable	Model	Hs	Mt	Training Dataset (2/3)			Test Dataset (1/3)			Full Dataset (3/3)		
				RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE
Spectral Bands	RF	1	11 [a]	0.04	0.99	0.03	0.09	0.91	0.07	0.06	0.96	0.04
		2	11 [a]	0.05	0.99	0.04	0.13	0.93	0.10	0.09	0.97	0.06
		3	11 [a]	0.05	0.99	0.03	0.12	0.93	0.09	0.08	0.97	0.05
NDVI		1	11 [a]	0.04	0.98	0.03	0.10	0.87	0.08	0.07	0.94	0.05
			2 [b]	0.10	0.89	0.08	0.20	0.51	0.16	0.14	0.76	0.11
		2	11 [a]	0.06	0.99	0.05	0.15	0.91	0.11	0.10	0.96	0.07
			2 [b]	0.10	0.96	0.08	0.21	0.81	0.16	0.15	0.91	0.11
		3	11 [a]	0.07	0.98	0.05	0.16	0.86	0.12	0.11	0.94	0.08
			2 [b]	0.12	0.93	0.09	0.25	0.68	0.19	0.17	0.84	0.13
GNDVI		1	11 [a]	0.05	0.97	0.04	0.12	0.83	0.09	0.08	0.93	0.05
			2 [b]	0.10	0.90	0.08	0.19	0.53	0.16	0.14	0.77	0.11
		2	11 [a]	0.06	0.98	0.05	0.15	0.90	0.12	0.10	0.96	0.07
			2 [b]	0.10	0.96	0.08	0.21	0.82	0.16	0.15	0.91	0.11
		3	11 [a]	0.07	0.97	0.06	0.17	0.84	0.14	0.12	0.93	0.08
			2 [b]	0.12	0.93	0.09	0.24	0.69	0.19	0.17	0.85	0.12
Spectral Bands	MLR	1	11 [a]	0.12	0.81	0.10	0.12	0.81	0.10	0.12	0.81	0.10
		2	11 [a]	0.17	0.88	0.13	0.17	0.88	0.14	0.17	0.88	0.14
		3	11 [a]	0.16	0.86	0.13	0.16	0.86	0.13	0.16	0.86	0.13
NDVI		1	11 [a]	0.14	0.77	0.11	0.14	0.77	0.11	0.14	0.77	0.11
			2 [b]	0.20	0.49	0.17	0.20	0.50	0.16	0.20	0.50	0.17
		2	11 [a]	0.19	0.86	0.15	0.19	0.86	0.15	0.19	0.86	0.15
			2 [b]	0.21	0.83	0.16	0.21	0.82	0.17	0.21	0.82	0.16
		3	11 [a]	0.21	0.77	0.17	0.21	0.76	0.16	0.21	0.76	0.17
			2 [b]	0.24	0.70	0.19	0.23	0.70	0.19	0.24	0.70	0.19
GNDVI		1	11 [a]	0.14	0.74	0.11	0.14	0.74	0.12	0.14	0.74	0.11
			2 [b]	0.20	0.51	0.16	0.20	0.52	0.16	0.20	0.51	0.16
		2	11 [a]	0.18	0.87	0.14	0.19	0.86	0.15	0.18	0.86	0.15
			2 [b]	0.21	0.82	0.17	0.21	0.82	0.17	0.21	0.82	0.17
		3	11 [a]	0.22	0.75	0.17	0.22	0.75	0.17	0.22	0.75	0.17
			2 [b]	0.24	0.70	0.19	0.23	0.70	0.19	0.24	0.70	0.19

Hs = Harvest; Mt = Months; RF = Random Forest regression; MLR = Multiple linear regression; [a]: satellite-imagery data from January, February, March, April, May, June, July, August, September, October and December were considered; [b]: satellite-imagery data from July and August were considered.

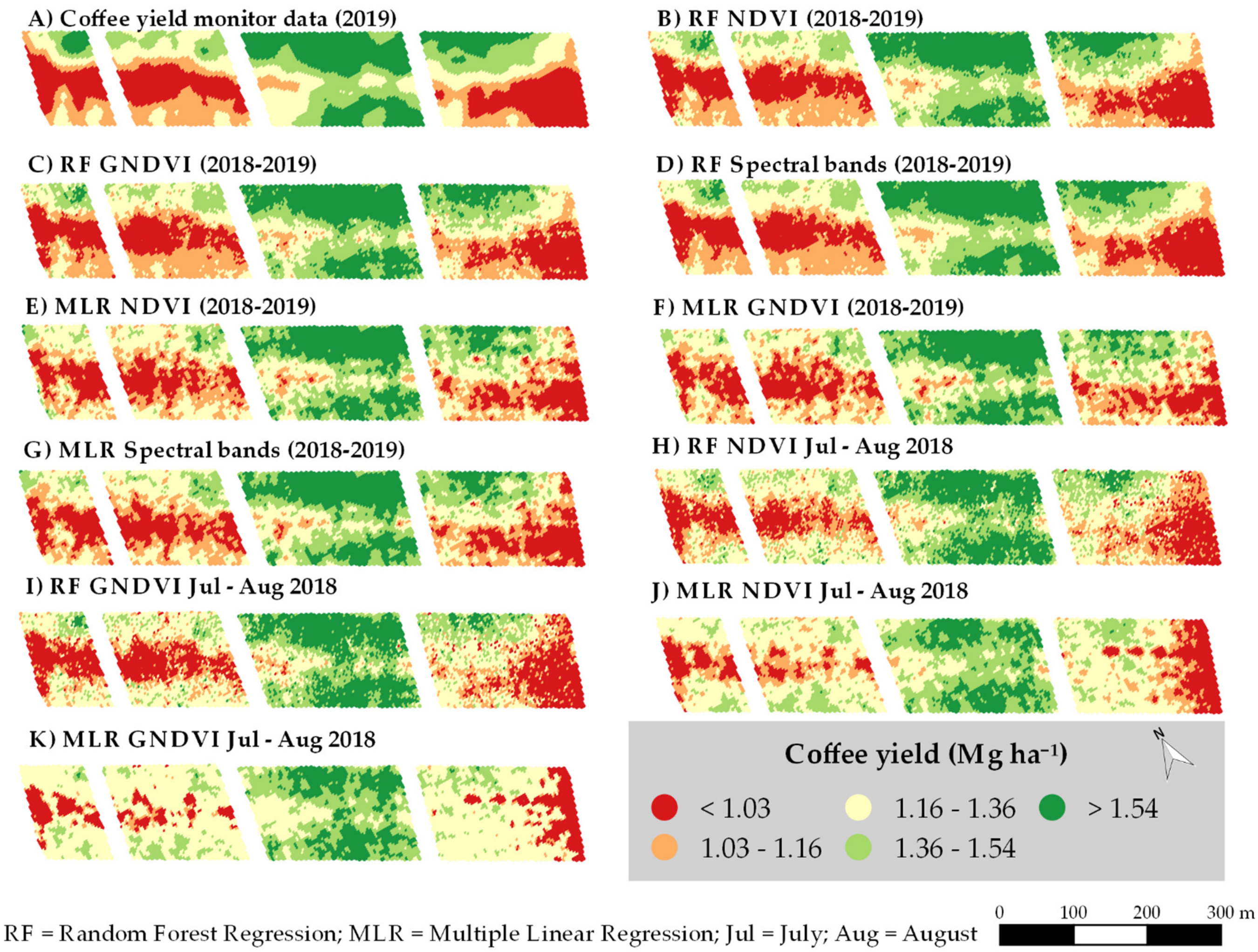

Figure 4. Coffee yield (Mg ha^{-1}) maps generate for harvest 1 (2018–2019) from (**A**) Coffee monitor data/2019; (**B**) RF regression model based on spectral bands/2018/2019, (**C**) RF regression based on NDVI/2018/2019, (**D**) RF regression based on GNDVI/2018/2019, (**E**) MLR based on spectral bands/2018/2019, (**F**) MLR based on NDVI/2018/2019, (**G**) MLR based on GNDVI/2018/2019, (**H**) RF regression based on NDVI obtained in July (jul) and August (ago)/2018, (**I**) RF regression based on GNDVI obtained from July and August/2018, (**J**) MLR based on NDVI obtained from July and August/2018; and (**K**) MLR based on GNDVI obtained from July and August/2018.

Considering Figures 4–6 it is inferred that RF regression models present smoother results when compared to MLR, regardless of the database. However, it can be seen that the larger the number of available variables (spectral bands), the smoother the results. The results from VIs and MLR are also worth using, but be aware of the computing power for processing data; the use of spectral bands to fit coffee yield-prediction models are of primary importance, as shown in other studies with different crops: carrot [53], sugarcane [34], corn [58] and soybean [58,59].

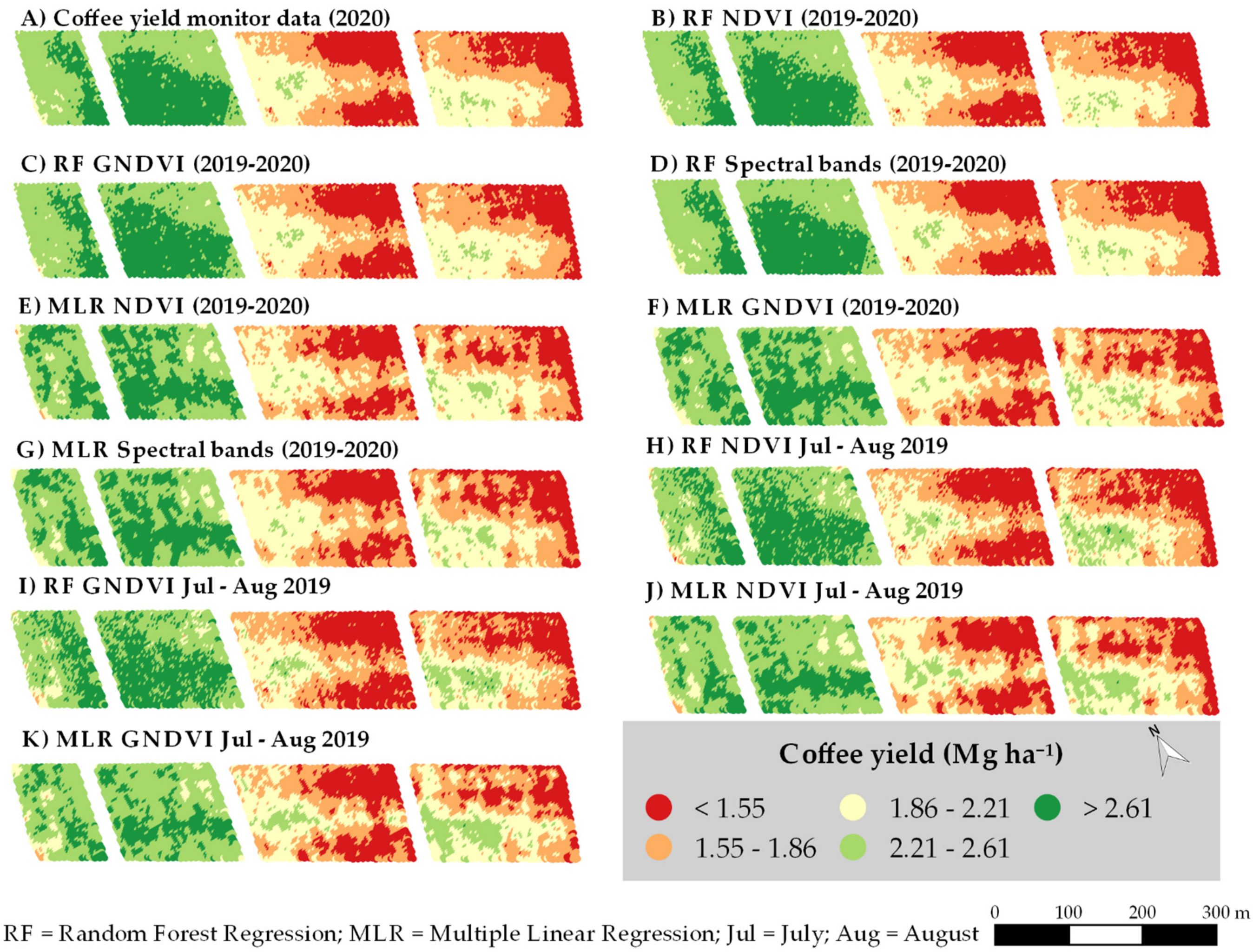

Figure 5. Coffee yield (Mg ha^{-1}) maps generate for harvest 2 (2019–2020) from (**A**) Coffee monitor data/2020; (**B**) RF regression model based on spectral bands/2019/2020, (**C**) RF regression based on NDVI/2019/2020, (**D**) RF regression based on GNDVI/2019/2020, (**E**) MLR based on spectral bands/2019/2020, (**F**) MLR based on NDVI/2019/2020, (**G**) MLR based on GNDVI/2019/2020, (**H**) RF regression based on NDVI obtained in July (jul) and August (ago)/2019, (**I**) RF regression based on GNDVI obtained from July and August/2019, (**J**) MLR based on NDVI obtained from July and August/2019; and (**K**) MLR based on GNDVI obtained from July and August/2019.

In addition, it is highlighted that regardless of the regression model used to fit predictive models, they all present similar data, allowing one to reproduce yield maps, in general, with similar spatial patterns, which can be used not only as a method to estimate yield, but also to indicate the spatial variability as high- and low-yielding zones.

The use of VIs in this work allowed us to identify potential management zones qualitatively as they present strong linear correlations with yield (Figure 3) due to their established importance for coffee physiology, since the images are acquired in the ideal crop phenological phase. Thus, NDVI/GNDVI data collection after harvesting (dry season—July and August—dormancy phase) allow us to improve management decisions regarding the crop to ensure it expresses their yield potential as much as possible. It is also possible to infer qualitatively about high- and low-yielding zones one year prior to the harvest season, valuable data that are able to be used by the decision maker to enhance crop management considering the crop variability.

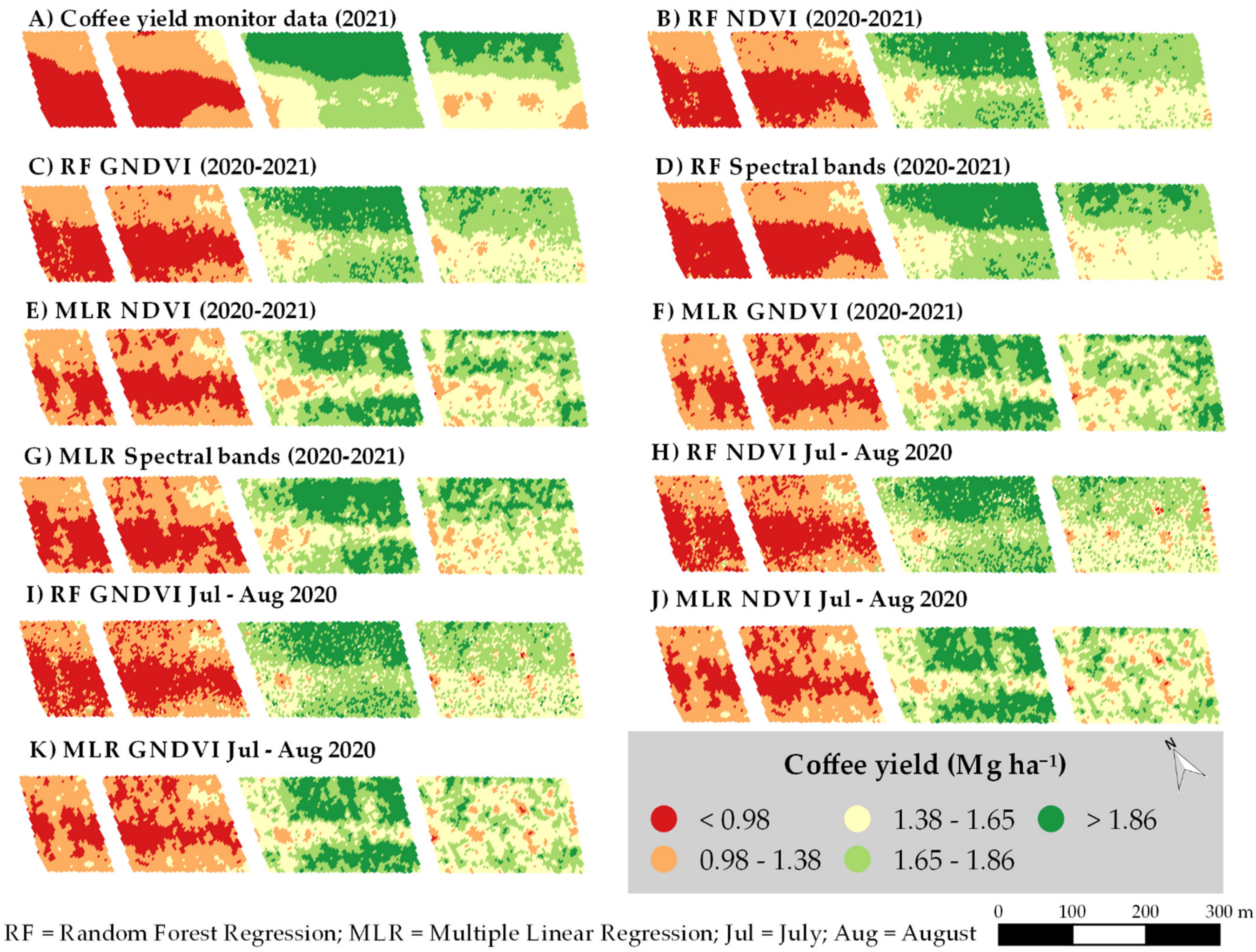

Figure 6. Coffee yield (Mg ha^{-1}) maps generate for harvest 2 (2020–2021) from (**A**) Coffee monitor data/2021; (**B**) RF regression model based on spectral bands/2020/2021, (**C**) RF regression based on NDVI/2020/2021, (**D**) RF regression based on GNDVI/2020/2021, (**E**) MLR based on spectral bands/2019/2020, (**F**) MLR based on NDVI/2020/2021, (**G**) MLR based on GNDVI/2020/2021, (**H**) RF regression based on NDVI obtained in July (jul) and August (ago)/2020, (**I**) RF regression based on GNDVI obtained from July and August/2020, (**J**) MLR based on NDVI obtained from July and August/2020; and (**K**) MLR based on GNDVI obtained from July and August/2020.

Note that the strategy described in this study, using NDVI/GNDVI data obtained during the dormancy phase of coffee plants that allow one to identify the qualitative yield-potential zones before the flowering occurs, can be critical to decision making in coffee production systems. Irrigation planning can be developed based on this information to match the crucial moment to enhance coffee quality [60,61]. Additionally, the optimal soil correctives and fertilizer application can be planned using zone prediction, following the principles of PA [62]. The high spatial density in which the qualitative zones can be predicted using this strategy can guide the prescription of coffee management across the season before blossoming and coffee-fruit development.

The dormancy phase is critical to the coffee plant, as it is waiting for the first rains to bloom [61]. Aware that blossoming occurs unevenly within the field, efforts should be made to optimize the planning and use of site-specific irrigation systems, since they can be improved by means of vegetative index data (NDVI/GNDVI) obtained in the dry season and related to yield.

Several studies have been conducted evaluating irrigation and coffee attributes qualitatively and quantitatively. For example, Rodrigues et al. [60] found that better coffee quality is found in irrigated areas, but the watering time is crucial. In addition, several morphological structures are also enhanced, which can be expressed in yield increment. Damatta et al. [63] also found that irrigated areas may produce coffee with better quality, but they highlighted that the timing of water provision can negatively affect the quality. Therefore, irrigation can improve the yield expression in terms of quality and quantity; if irrigation occurs at the correct time, an additional data layer (yield potential) could also be used to guarantee that yield expression follows the principles of PA.

Fertilization management can also be guided by the yield map [29]. Therefore, fertilizer application should be distributed unevenly according to yield potential based on the NDVI/GNDVI map.

In terms of quantitative data, it was demonstrated in this work that the use of field data (harvester yield data) with satellite images applying regression models are suitable for estimating coffee yield. Future approaches can explore the use of these models to extrapolate the yield estimate to nearby areas, as well as evaluate the potential of the technique for applying a model to the next year, as well as testing other models for prediction such as SVR, NN and other vegetation indices.

4. Conclusions

It was possible to observe with the set of PlanetScope orbital images and the yield data during three harvests that there is a direct correlation between VIs (NDVI/GNDVI) and yield zones one year before harvest, especially in the months of July and August (post-harvest and in the dormancy phase of the plant). The use of the proposed strategy to delimitate high- and low-yield zones can be a critical guide for crop monitoring and management practices during the season.

Using regression models (RF and MLR) it was possible to estimate the coffee yield. The RF regression models showed the highest R^2 (0.93) values compared to the MLR (0.88) in the same period. Comparing the results based on the time series (11 months) and only on the two best months (July and August), it is noted that the fewer variables used, the lower the accuracy, independent of the dataset and regression model. However, when it is not possible to obtain a set of annual images, the results showed that it is possible to estimate yield with images during the dormancy phase of the plant one year before harvest. This offers an alternative cost-efficient strategy that enable producers to monitor yield and estimate profit, and especially guide management practices by the premises of precision agriculture, taking into account the temporal and spatial variability of the field.

In addition, it is noteworthy that regardless of the regression model used to adjust the predictive models, they all present similar data, allowing one to reproduce yield maps, in general, with similar spatial patterns, which can be used not only as a method to estimate yield, but also to indicate spatial variability with zones of high and low yield, observing the behavior of productive alternation in the area and indicating the presence of biennial yield.

Author Contributions: Conceptualization, M.M. and J.P.M.; methodology, M.M., J.P.M., M.C.F.W. and J.V.M.N.; software, M.M. and M.C.F.W.; formal analysis, M.M., M.C.F.W. and R.C.F.; writing—original draft preparation, M.M., J.P.M. and M.C.F.W.; writing—review and editing, M.M., J.P.M., M.C.F.W., R.C.F. and J.V.M.N. All authors have read and agreed to the published version of the manuscript.

Funding: Mauricio Martello was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001. Ricardo Canal Filho was funded by the National Council for Scientific and Technological Development (CNPq), Grant No. 830707/1999-9.

Data Availability Statement: Not applicable.

Acknowledgments: To the Guima Café Group for allowing the use of their plantations as experimental areas and providing the people and machinery, Terrena Agronegócios for supporting the research.

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

References

1. Santana, L.S.; Ferraz, G.A.E.S.; Teodoro, A.J.d.S.; Santana, M.S.; Rossi, G.; Palchetti, E. Advances in Precision Coffee Growing Research: A Bibliometric Review. *Agronomy* **2021**, *11*, 1557. [CrossRef]
2. Santinato, F. Inovações Tecnológicas Para Cafeicultura de Precisão. Ph.D. Thesis, School of Agricultural and Veterinarian Studies, São Paulo State University, Jaboticabal, Brazil, 2016.
3. Chemura, A.; Mutanga, O.; Odindi, J.; Kutywayo, D. Mapping Spatial Variability of Foliar Nitrogen in Coffee (*Coffea arabica* L.) Plantations with Multispectral Sentinel-2 MSI Data. *ISPRS J. Photogramm. Remote Sens.* **2018**, *138*, 1–11. [CrossRef]
4. Pham, Y.; Reardon-Smith, K.; Mushtaq, S.; Cockfield, G. The Impact of Climate Change and Variability on Coffee Production: A Systematic Review. *Clim. Change* **2019**, *156*, 609–630. [CrossRef]
5. Marin, D.B.; Ferraz, G.A.e.S.; Guimarães, P.H.S.; Schwerz, F.; Santana, L.S.; Barbosa, B.D.S.; Barata, R.A.P.; Faria, R.d.O.; Dias, J.E.L.; Conti, L.; et al. Remotely Piloted Aircraft and Random Forest in the Evaluation of the Spatial Variability of Foliar Nitrogen in Coffee Crop. *Remote Sens.* **2021**, *13*, 1471. [CrossRef]
6. Bazame, H.C.; Molin, J.P.; Althoff, D.; Martello, M. Detection, Classification, and Mapping of Coffee Fruits during Harvest with Computer Vision. *Comput. Electron. Agric.* **2021**, *183*, 106066. [CrossRef]
7. Mulla, D.J. Twenty Five Years of Remote Sensing in Precision Agriculture: Key Advances and Remaining Knowledge Gaps. *Biosyst. Eng.* **2013**, *114*, 358–371. [CrossRef]
8. Kayad, A.; Sozzi, M.; Gatto, S.; Marinello, F.; Pirotti, F. Monitoring Within-Field Variability of Corn Yield Using Sentinel-2 and Machine Learning Techniques. *Remote Sens.* **2019**, *11*, 2873. [CrossRef]
9. Damian, J.M.; Pias, O.H.d.C.; Cherubin, M.R.; da Fonseca, A.Z.; Fornari, E.Z.; Santi, A.L. Applying the NDVI from Satellite Images in Delimiting Management Zones for Annual Crops. *Sci. Agric.* **2020**, *77*, e20180055. [CrossRef]
10. El-Ghany, N.M.A.; El-Aziz, S.E.A.; Marei, S.S. A Review: Application of Remote Sensing as a Promising Strategy for Insect Pests and Diseases Management. *Environ. Sci. Pollut. Res.* **2020**, *27*, 33503–33515. [CrossRef]
11. Karthikeyan, L.; Chawla, I.; Mishra, A.K. A Review of Remote Sensing Applications in Agriculture for Food Security: Crop Growth and Yield, Irrigation, and Crop Losses. *J. Hydrol.* **2020**, *586*, 124905. [CrossRef]
12. Saraiva, M.; Protas, É.; Salgado, M.; Souza, C. Automatic Mapping of Center Pivot Irrigation Systems from Satellite Images Using Deep Learning. *Remote Sens.* **2020**, *12*, 558. [CrossRef]
13. Fabbri, C.; Mancini, M.; Marta, A.D.; Orlandini, S.; Napoli, M. Integrating Satellite Data with a Nitrogen Nutrition Curve for Precision Top-Dress Fertilization of Durum Wheat. *Eur. J. Agron.* **2020**, *120*, 126148. [CrossRef]
14. Lobell, D.B.; Thau, D.; Seifert, C.; Engle, E.; Little, B. A Scalable Satellite-Based Crop Yield Mapper. *Remote Sens. Environ.* **2015**, *164*, 324–333. [CrossRef]
15. Colaço, A.F.; Bramley, R.G.V. Do Crop Sensors Promote Improved Nitrogen Management in Grain Crops? *Field Crops Res.* **2018**, *218*, 126–140. [CrossRef]
16. Bramley, R.G.V.; Ouzman, J.; Gobbett, D.L. Regional Scale Application of the Precision Agriculture Thought Process to Promote Improved Fertilizer Management in the Australian Sugar Industry. *Precis. Agric.* **2019**, *20*, 362–378. [CrossRef]
17. Luo, X.; Ye, Z.; Xu, H.; Zhang, D.; Bai, S.; Ying, Y. Robustness Improvement of NIR-Based Determination of Soluble Solids in Apple Fruit by Local Calibration. *Postharvest Biol. Technol.* **2018**, *139*, 82–90. [CrossRef]
18. Lawes, R.A.; Oliver, Y.M.; Huth, N.I. Optimal Nitrogen Rate Can Be Predicted Using Average Yield and Estimates of Soil Water and Leaf Nitrogen with Infield Experimentation. *Agron. J.* **2019**, *111*, 1155–1164. [CrossRef]
19. Padarian, J.; Minasny, B.; McBratney, A.B. Transfer Learning to Localise a Continental Soil Vis-NIR Calibration Model. *Geoderma* **2019**, *340*, 279–288. [CrossRef]
20. Vega, A.; Córdoba, M.; Castro-Franco, M.; Balzarini, M. Protocol for Automating Error Removal from Yield Maps. *Precis. Agric.* **2019**, *20*, 1030–1044. [CrossRef]
21. Jeffries, G.R.; Griffin, T.S.; Fleisher, D.H.; Naumova, E.N.; Koch, M.; Wardlow, B.D. Mapping Sub-Field Maize Yields in Nebraska, USA by Combining Remote Sensing Imagery, Crop Simulation Models, and Machine Learning. *Precis. Agric.* **2020**, *21*, 678–694. [CrossRef]
22. Momin, M.A.; Grift, T.E.; Valente, D.S.; Hansen, A.C. Sugarcane Yield Mapping Based on Vehicle Tracking. *Precis. Agric.* **2019**, *20*, 896–910. [CrossRef]
23. De Silva, F.M.; de Souza, Z.M.; de Figueiredo, C.A.P.; Vieira, L.H.S.; de Oliveira, E. Variabilidade Espacial de Atributos Químicos e Produtividade Da Cultura Do Café Em Duas Safras Agrícolas. *Ciência Agrotecnologia* **2008**, *32*, 231–241. [CrossRef]
24. Ferraz, G.; da Silva, F.; Carvalho, L.C.C.; Alves, M.D.C.; Franco, B.C. Spatial and Temporal Variability of Phosphorous, Potassium and of the Yield of a Coffee Field. *Eng. Agríc. Jaboticabal* **2012**, *32*, 140–150. [CrossRef]
25. Carvalho, L.C.C.; da Silva, F.M.; Ferraz, G.A.E.S.; Stracieri, J.; Ferraz, P.F.P.; Ambrosano, L. Geostatistical Analysis of Arabic Coffee Yield in Two Crop Seasons. *Rev. Bras. Eng. Agric. Ambient.* **2017**, *21*, 410–414. [CrossRef]
26. Sartori, S.; Fava, J.F.M.; Domingues, E.L.; Ribeiro Filho, A.C.; Shiraisi, L.E. *Mapping the Spatial Variability of Coffee Yield with Mechanical Harvester*; American Society of Agricultural and Biological Engineers: Saint Joseph, MI, USA, 2013.
27. Martello, M.; Molin, J.P.; Bazame, H.C. Obtaining and Validating High-Density Coffee Yield Data. *Horticulturae* **2022**, *8*, 421. [CrossRef]

28. Rahnemoonfar, M.; Sheppard, C. Deep Count: Fruit Counting Based on Deep Simulated Learning. *Sensors* **2017**, *17*, 905. [CrossRef]
29. Molin, J.P.; Motomiya, A.V.d.A.; Frasson, F.R.; Faulin, G.D.C.; Tosta, W. Test Procedure for Variable Rate Fertilizer on Coffee. *Acta Sci. Agron.* **2010**, *32*, 569–575. [CrossRef]
30. Bernardes, T.; Moreira, M.A.; Adami, M.; Giarolla, A.; Rudorff, B.F.T. Monitoring Biennial Bearing Effect on Coffee Yield Using MODIS Remote Sensing Imagery. *Remote Sens.* **2012**, *4*, 2492. [CrossRef]
31. Nogueira, S.M.C.; Moreira, M.A.; Volpato, M.M.L. Relationship between Coffee Crop Yield and Vegetation Indexes Derived from Oli/Landsat-8 Sensor Data with and without Topographic Correction. *Eng. Agric.* **2018**, *38*, 387–394. [CrossRef]
32. Thao, N.T.T.; Khoi, D.N.; Denis, A.; Viet, L.V.; Wellens, J.; Tychon, B. Early Prediction of Coffee Yield in the Central Highlands of Vietnam Using a Statistical Approach and Satellite Remote Sensing Vegetation Biophysical Variables. *Remote Sens.* **2022**, *14*, 2975. [CrossRef]
33. Silva, P.A.D.A.; Alves, M.d.C.; da Silva, F.M.; Figueiredo, V.C. Coffee Yield Estimation by Landsat-8 Imagery Considering Shading Effects of Planting Row's Orientation in Center Pivot. *Remote Sens. Appl. Soc. Environ.* **2021**, *24*, 100613. [CrossRef]
34. Canata, T.F.; Wei, M.C.F.; Maldaner, L.F.; Molin, J.P. Sugarcane Yield Mapping Using High-Resolution Imagery Data and Machine Learning Technique. *Remote Sens.* **2021**, *13*, 232. [CrossRef]
35. Hunt, D.A.; Tabor, K.; Hewson, J.H.; Wood, M.A.; Reymondin, L.; Koenig, K.; Schmitt-Harsh, M.; Follett, F. Review of Remote Sensing Methods to Map Coffee Production Systems. *Remote Sens.* **2020**, *12*, 2041. [CrossRef]
36. Jeong, J.H.; Resop, J.P.; Mueller, N.D.; Fleisher, D.H.; Yun, K.; Butler, E.E.; Timlin, D.J.; Shim, K.M.; Gerber, J.S.; Reddy, V.R.; et al. Random Forests for Global and Regional Crop Yield Predictions. *PLoS ONE* **2016**, *11*, e0156571. [CrossRef]
37. Hochachka, W.M.; Caruana, R.; Fink, D.; Munson, A.; Riedewald, M.; Sorokina, D.; Kelling, S. Data-Mining Discovery of Pattern and Process in Ecological Systems. *J. Wildl. Manag.* **2007**, *71*, 2427. [CrossRef]
38. Alvares, C.A.; Stape, J.L.; Sentelhas, P.C.; de Moraes Gonçalves, J.L.; Sparovek, G. Köppen's Climate Classification Map for Brazil. *Meteorol. Z.* **2013**, *22*, 711–728. [CrossRef]
39. INMET. *Instituto Nacional de Meteorologia: Brazil Climate Normals 1991–2020*; INMET: Brasília, Brazil, 2022. Available online: https://portal.inmet.gov.br/uploads/normais/NORMAISCLIMATOLOGICAS.pdf (accessed on 15 July 2022).
40. Maldaner, L.F.; Molin, J.P. Data Processing within Rows for Sugarcane Yield Mapping. *Sci. Agric.* **2020**, *77*, e20180391. [CrossRef]
41. Minasny, B.; McBratney, A.B.; Whelan, B.M. *VESPER, Version 1.62*; University of Sydney: Sydney, Australia, 2006.
42. Planet Team. Planet Application Program Interface: In Space for Life on Earth. *San Francisco, CA, USA.* 2017. Available online: https://api.planet.com (accessed on 15 July 2022).
43. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. *Monitoring Vegetation Systems in the Great Plains with ERTS*; NASA Special Publication; NASA: Washington, DC, USA, 1974; p. 24.
44. Rodríguez-López, L.; Duran-Llacer, I.; González-Rodríguez, L.; Abarca-Del-Rio, R.; Cárdenas, R.; Parra, O.; Martínez-Retureta, R.; Urrutia, R. Spectral analysis using LANDSAT images to monitor the chlorophyll-a concentration in Lake Laja in Chile. *Ecol. Inform.* **2020**, *60*, 101183. [CrossRef]
45. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a Green Channel in Remote Sensing of Global Vegetation from EOS- MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [CrossRef]
46. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
47. Liaw, A.; Wiener, M. Classification and Regression by RandomForest. *R News* **2002**, *2*, 18–22.
48. Olive, D.J. Multivariate Linear Regression. In *Linear Regression*; Olive, D.J., Ed.; Springer: Chan, Switzerland, 2017; pp. 17–83.
49. QGIS Development Team. QGIS Geographic Information System. *Open Source Geospatial Foundation Project.* 2014. Available online: http://Qgis.Osgeo.Org (accessed on 15 July 2022).
50. De Oliveira, R.R.; Cesarino, I.; Mazzafera, P.; Dornelas, M.C. Flower Development in Coffea Arabica L.: New Insights into MADS-Box Genes. *Plant Reprod.* **2014**, *27*, 79–94. [CrossRef]
51. De Camargo, Â.P.; de Camargo, M.B.P. Definition and Outline for the Phenological Phases of Arabic Coffee under Brazilian Tropical Conditions. *Bragantia* **2001**, *60*, 65–68. [CrossRef]
52. Lima, A.A.; Santos, I.S.; Torres, M.E.L.; Cardon, C.H.; Caldeira, C.F.; Lima, R.R.; Davies, W.J.; Dodd, I.C.; Chalfun-Junior, A. Drought and Re-Watering Modify Ethylene Production and Sensitivity, and Are Associated with Coffee Anthesis. *Environ. Exp. Bot.* **2021**, *181*, 104289. [CrossRef]
53. Rena, A.B.; Maestri, M. *Fisiologia Do Cafeeiro. Cultura Do Cafeeiro: Fatores Que Afetam a Produtividad*; Associação Brasileira para Pesquisa da Potassa e do Fosfato: Piracicaba, Brazil, 1986.
54. Pereira, S.P.; Bartholo, G.F.; Baliza, D.P.; Sobreira, F.M.; Guimarães, R.J. Growth, Yield and Bienniality of Coffee Plants According to Cultivation Spacing | Crescimento, Produtividade e Bienalidade Do Cafeeiro Em Função Do Espaçamento de Cultivo. *Pesqui. Agropecu. Bras.* **2011**, *46*, 152–160. [CrossRef]
55. De Gaspari-Pezzopane, C.; Medina Filho, H.P.; Bordignon, R.; Siqueira, W.J.; Ambrósio, L.A.; Mazzafera, P. Environmental Influences on the Intrinsic Outturn of Coffee. *Bragantia* **2005**, *64*, 39–50.
56. Wei, M.C.F.; Maldaner, L.F.; Ottoni, P.M.N.; Molin, J.P. Carrot Yield Mapping: A Precision Agriculture Approach Based on Machine Learning. *AI* **2020**, *1*, 15. [CrossRef]
57. Angnes, G.; Martello, M.; Faulin, G.D.C.; Molin, J.P.; Romanelli, T.L. Energy Efficiency of Variable Rate Fertilizer Application in Coffee Production in Brazil. *AgriEngineering* **2021**, *3*, 51. [CrossRef]

58. Skakun, S.; Brown, M.G.L.; Roger, J.C.; Vermote, E. Capturing Corn and Soybean Yield Variability at Field Scale Using Very High Spatial Resolution Satellite Data. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Virtual, 26 September–2 October 2020.
59. Gava, R.; Santana, D.C.; Cotrim, M.F.; Rossi, F.S.; Teodoro, L.P.R.; da Silva Junior, C.A.; Teodoro, P.E. Soybean Cultivars Identification Using Remotely Sensed Image and Machine Learning Models. *Sustainability* **2022**, *14*, 7125. [CrossRef]
60. Rodrigues, W.N.; Brinate, S.V.B.; Martins, L.D.; Colodetti, T.V.; Tomaz, M.A. Genetic Variability and Expression of Agro-Morphological Traits among Genotypes of Coffea Arabica Being Promoted by Supplementary Irrigation. *Genet. Mol. Res.* **2017**, *16*, gmr16029563. [CrossRef]
61. Miranda, F.R.; Drumond, L.C.D.; Ronchi, C.P. Synchronizing Coffee Blossoming and Fruit Ripening in Irrigated Crops of the Brazilian Cerrado Mineiro Region. *Aust. J. Crop Sci.* **2020**, *14*, 605–613. [CrossRef]
62. International Society of Precision Agriculture (ISPA). Precision Ag Definition. Available online: https://www.ispag.org/about/definition (accessed on 15 July 2022).
63. DaMatta, F.M.; Ronchi, C.P.; Maestri, M.; Barros, R.S. Ecophysiology of Coffee Growth and Production. *Braz. J. Plant Physiol.* **2007**, *19*, 485–510. [CrossRef]