Article

Early Prediction of Coffee Yield in the Central Highlands of Vietnam Using a Statistical Approach and Satellite Remote Sensing Vegetation Biophysical Variables

Nguyen Thi Thanh Thao [1,2,*], Dao Nguyen Khoi [3,4], Antoine Denis [1], Luong Van Viet [2], Joost Wellens [1] and Bernard Tychon [1]

1 Spheres Research Unit, Water, Environment and Development Laboratory, Environmental Sciences and Management Department, Arlon Campus Environment, University of Liège, 185 Avenue de Longwy, 6700 Arlon, Belgium; antoine.denis@uliege.be (A.D.); joost.wellens@uliege.be (J.W.); bernard.tychon@uliege.be (B.T.)
2 Institute of Environmental Science, Engineering and Management, Industrial University of Ho Chi Minh City, Ho Chi Minh City 700000, Vietnam; luongvanviet@iuh.edu.vn
3 Faculty of Environment, University of Science, Ho Chi Minh City 700000, Vietnam; dnkhoi@hcmus.edu.vn
4 Vietnam National University Ho Chi Minh City, Linh Trung Ward, Thu Duc District, Ho Chi Minh City 700000, Vietnam
* Correspondence: thao.smile06@gmail.com or nguyenthithanhthao@iuh.edu.vn

Abstract: Given the present climate change context, accurate and timely coffee yield prediction is critical to all farmers who work in the coffee industry worldwide. The aim of this study is to develop and assess a coffee yield forecasting method at the regional scale in Dak Lak province in the central highlands of Vietnam using the Crop Growth Monitoring System Statistical Tool (CGMSstatTool—CST) software and vegetation biophysical variables (NDVI, LAI, and FAPAR) derived from satellite remote sensing (SPOT-VEGETATION and PROBA-V). There has been no research to date applying this approach to this specific crop, which is the main contribution of this study. The findings of this research reveal that the elaboration of multiple linear regression models based on a combination of information from satellite-derived vegetation biophysical variables (LAI, NDVI, and FAPAR) corresponding to the first six months of the years 2000–2019 resulted in coffee yield forecast models presenting satisfactory accuracy (Adj.R^2 = 64 to 69%, RMSEp = 0.155 to 0.158 ton/ha and MAPE = 3.9 to 4.7%). These results demonstrate that the CST may efficiently predict coffee yields on a regional scale by using only satellite-derived vegetation biophysical variables. This study findings are likely to aid local governments and decision makers in precisely forecasting coffee production early and promptly, as well as in recommending relevant local agricultural policies.

Keywords: coffee yield forecast; remote sensing vegetation biophysical variables; early prediction; CGMSstatTool; Vietnam; LAI; NDVI; FAPAR

1. Introduction

Coffee is one of the most crucial agricultural products in the global market, playing a significant part in the economy of several developing countries in equatorial and subequatorial regions (Africa, America, and Asia) [1–3]. Currently, two coffee bean species, *Coffee arabica* L. (Arabica coffee) and *C. canephora* Pierre ex A. Froehner (Robusta coffee), account for 99% of coffee production in the global coffee trade [3,4]. Coffee is grown in approximately 80 tropical countries and contributes to the economic base of many of these countries. In addition, about 25 million farmer families produce coffee worldwide, with most being smallholders and families whose source of revenue largely depends on this crop [4]. Coffee is a climate-sensitive perennial plant likely to be highly influenced by changes in climate. Increasing climate variability may lead to coffee yield decrease and coffee area damage and threaten coffee production in producing areas worldwide [5,6].

Extreme weather events, such as severe droughts or excess precipitation, in these parts of the world associated with the El Nino Southern Oscillation (ENSO) significantly influence coffee production in the global market [1,3].

Vietnam is the world's second largest exporter of coffee beans, with a total average annual coffee production of 25.73 million 60-kg bags (1544 million metric tons) during the period from 2011 to 2021, accounting for roughly 20 percent of the world's coffee production [7,8], and making it the largest global producer of robusta coffee [3]. The central highlands area is one of the most critical regions for Vietnam's economy because it is the largest producer of coffee beans in Vietnam [3,9]—mainly robusta coffee [3]. Robusta coffee yield is influenced by the interaction between precipitation, temperature, and phenological stages. Robusta coffee reacts better to rising temperatures than arabica coffee [6,10,11] and is considered more resistant to climate change than other coffee species [11]. In recent years, increasing temperatures and variability in precipitation in Vietnam's subregions were associated with the El Nino Southern Oscillation [12,13]. Weather data indicated lower precipitation and higher average temperatures than mean conditions in Vietnam's major coffee-growing areas for the first five months of the calendar year 2020, causing lower yields and reduced production [14]. Therefore, it is necessary to provide decision makers with support tools enabling forecasting of coffee yield and production in order to facilitate the development of management strategies and of economic evaluation of coffee production for the various stakeholders of the coffee industry, from smallholder farmers to governmental authorities.

Models to simulate and predict coffee production have been developed in several studies. Gutierrez et al. (1998) [15] developed a model to simulate the vegetative growth process of arabica coffee as age–mass-structured populations of stem, root, and leaves and enabled branch-level computation of leaf area index at any stage of coffee development. Furthermore, Rodríguez et al. (2011) [16] built a model to simulate the phenology, growth, and development of coffee plants based on the physiologically based models of Gutierrez et al. (1998) [15]. The model inputs consisted of soil parameters (e.g., nitrogen and water) and daily meteorological data. This model easily incorporated other coffee varieties in different ecological zones where coffee is cultivated [16]. The model was successfully calibrated for the Colombian and Brazilian regions, two areas with differing climates and flower phenology (subtropical and equatorial) [17]. Van Oijen et al. (2010) also developed a simple dynamic model of coffee agroforestry systems that models the physiology of vegetative and reproductive growth of coffee plants and their response to different cultivating conditions [18]. The strengths of this model are its ease of use, its speed, and that it can be run under changing climatic conditions. Growing conditions such as weather conditions (temperature, rain, light, humidity, and wind), soil conditions (initial organic matter and nitrogen content, water balance, etc.), tree management (choice of species, density, etc.), and coffee management (rotation length, fertilization, and pruning regime) are addressed by the model as inputs. Rahn et al. (2018) [19] applied this model to two sites in East Africa with different climates. It was also calibrated and modified successfully in two different coffee-growing sites in Costa Rica and Nicaragua by Ovalle-Rivera et al. (2020) [20]. In addition, Vezy et al. (2020) [17] designed a DynAC of model to incorporate a plant-scale reproductive phenology formalism of the Rodríguez et al. (2011) model [16]. It was based on canopy temperature, with distinct submodules to obtain suitable adjustment of coffee and shade tree management, density, and tree species, as in the model of Van Oijen et al. (2010) [18] (i.e., canopy temperature-dependent phenology and the submodules for agroforestry system management). Kouadio et al. (2021) also successfully tested a process-based model using satellite remote sensing data (LAI) and model-based gridded climate data for predicting robusta coffee yield in the central highlands of Vietnam [3]. Kouadio et al. (2021) indicated that one of the limitations they encountered was the unavailability of distinct production statistics for arabica and robusta coffee. Aside from the process-based model [3], Kouadio et al. (2018) developed a model based on an artificial intelligence approach using soil fertility properties to predict robusta coffee yield in the Lam Dong

province of Vietnam [21]. Furthermore, Molina et al. (2018) successfully calibrated the Aquacrop model to predict arabica coffee yield in Colombia [22]. The input databases for this application contained parameters associated with climatic variables, soil, crops, and management practices.

Most of the aforementioned models simulate coffee yields by accounting for the growing conditions. These models mainly depend on the data collected in the field, as well as observed weather data. The accuracy of such models also depends on accurate descriptions of crop management practices (e.g., crop variety, sowing date, fertilization, and irrigation), although collecting such data in a sufficiently accurate manner is difficult at the regional scale [23]. Furthermore, several years of experimental data are necessary to train and calibrate models to the local environmental conditions for these crop models, and when they are applied in other regions, they have to be recalibrated [23].

Due to the limitations of these models, statistical models, such as multiple linear regression, have been widely utilized to link crop yields to climate variables [24,25] or even intermediate output variables from process-based crop models [26]. Despite not being directly based on the mechanisms of plant growth, statistical models can effectively predict crop production [23]. The main benefits of statistical models are their limited dependence on field calibration data and their clear assessment of model uncertainties [27]. Statistical models typically perform better as the availability and quality of observable data improve [23].

Among the various tools and methods that enable the development of statistically based models, the "Crop Growth Monitoring System Statistical Tool" (CGMSstatTool–CGMS statistical tool—CST) [28] is independent software for forecasting crop yield based on initial indicator databases derived from crop models, climate data, or remote sensing data [29]. The CST was developed by the MARS (Monitoring Agriculture with Remote Sensing) unit of the European Union (EU) Joint Research Centre (JRC) to support the development and selection of crop yield forecast models in order to assist national or subnational crop yield forecasting activities [28]. The CST plays a crucial role in scientific decision making in the EU agricultural economy [28]. The CST has been effectively applied to growth monitoring and yield forecasting of some main crops in northeast China [30]. Fall et al. (2021) also used the CST to predict millet yield at a regional scale in Senegal with input data containing weather data combined with variables derived from remote sensing indicators (NDVI) [23]. The CST enables simple examination of data quality, analysis of crop yield time trend, and construction of crop yield forecasting models through three methods: (1) multivariate regression analysis; (2) scenario analysis, which is a method of forecasting that looks for the previous years that are most similar to the current year based on a set of indicators and combines their yields [23]; and (3) the moving average analysis model, which is based on the average yields of the most recent years preceding the target year. The CST calculates a number of statistics that facilitate the choice of the best crop yield forecast model for a given region and time of prediction. Another advantage of the CST is its ability to rapidly test multiple models [23,28].

With the development of satellite imagery, agricultural monitoring systems have used indices derived from the spectral reflectance of vegetation to provide timely and concise information about seasonal vegetative growth [31]. Remote-sensing-derived vegetation indices (e.g., the normalized difference vegetation index, NDVI) and biophysical variables (e.g., the fraction of absorbed photosynthetically active radiation (FAPAR) and the leaf area index (LAI)) can be used to predict crop yield, either directly or indirectly [32,33]. In addition, remote sensing vegetation variables enable estimation of crop growth variability to quantify the relative development and health conditions of crops [34]. Such vegetation indices and biophysical variables are the most common satellite products utilized for these purposes [31]. At the national and regional levels, satellite systems can contribute effectively to early warning of crop stress during the growing period and in forecasting harvest yields [31,35]. In a study of the largest coffee-exporting province in Brazil using a dataset covering the 2002–2009 period, Bernardes et al. (2012) [36] observed correlations between

variations in the yield of coffee plots and variations in MODIS-derived EVI and NDVI vegetation indices computed from pure coffee crops (250 m pixels) overlapping the same coffee plots. The vegetation index metrics best correlated to yield were the amplitude and the minimum values over the growing season. The best correlations were obtained between the variation in yield and variation in vegetation indices of the previous year (R^2 = 0.55). In another study, Nogueira (2018) [37] evaluated the relationships between coffee productivity of some coffee plantations in Brazil and values of NDVI, SAVI, and NDWI vegetation indices derived from LANDSAT-8-OLI sensors for different coffee phenological phases. They concluded that the best phenological phases of coffee to determine coffee productivity from spectral indices were the stages of dormancy and flowering. The results also indicated that the NDVI was the best index to estimate the productivity of coffee trees, with the coefficient of determination (R^2) ranging from 0.58 to 0.90.

The objective of this research is to develop and assess a coffee yield forecasting method at the regional scale in Dak Lak province in the central highlands of Vietnam using Crop Growth Monitoring System Statistical Tool (CGMSstatTool—CST) software and vegetation biophysical variables (NDVI, LAI, and FAPAR) derived from satellite remote sensing (SPOT-VEGETATION and PROBA-V).

The findings of this study are expected to assist local governments and decision makers in accurately forecasting coffee yields early and in a timely manner, as well as in recommending appropriate strategies for local agriculture.
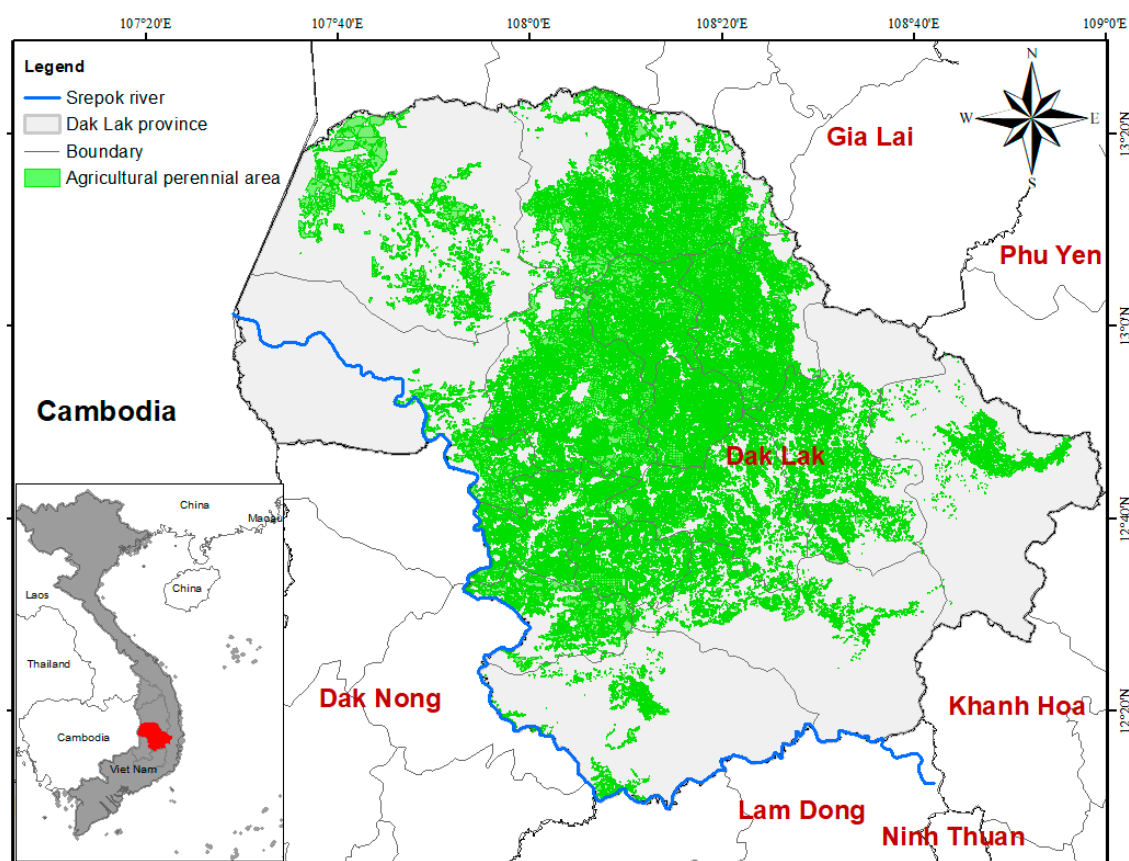
2. Method

2.1. Study Area

The study was carried out in Dak Lak province, located in the central highlands of Vietnam in the Lower Mekong River Basin. The total area is 13,125 km², and in 2019, the population of Dak Lak province was 2.127 million people. Currently, Dak Lak province includes Buon Ma Thuot city, Buon Ho town, and 13 districts.

In Dak Lak province, agriculture is the main source of local livelihoods. The area's geographic coordinates are from 107° to 109° east longitude and from 12° to 13° north latitude (Figure 1), with an average elevation range of 400–800 m. Dak Lak province is dominated by a humid tropical climate. Generally, the area climate varies depending on the altitude: below 300 m, it is hot all year round, in the range of 400–800 m it is hot and humid, and it is cool at altitudes above 800 m [38]. There are two distinct seasons in Dak Lak province: a rainy season from May to October, with approximately 80–85% of annual rainfall, and a dry season from November to April, which is generally dry and sunny (15–20% of annual rainfall). Dak Lak province is an agricultural area with perennial crops such as coffee, pepper, cashew, and fruits, which play an important part in its economy. The region also produces annual crops such as rice, maize, sweet potato, vegetables, sugarcane, groundnut, and soybean [39]. Dak Lak has 209,955 ha of coffee area, accounting for nearly 31% of the country's coffee area [40]. In Dak Lak province, coffee exports represent 86% of total agricultural exports and more than 60% of the total yearly provincial income. In addition, coffee production employs more than 300,000 direct workers and more than 100,000 indirect workers [41]. The vast majority of coffee trees are part of coffee tree plantations, where coffee trees are the main vegetation story. Irrigation has been applied one to four times per year since 2008 across robusta coffee crops in Dak Lak province (on average, 1345 L/tree/year, i.e., 148 mm/year). Irrigation quantities vary based on rainfall patterns during the coffee growing season [42].

The general methodological workflow followed in the present research is presented in Figure 2 and further detailed below.

Figure 1. Dak Lak province with the agricultural perennial planted area indicated in green. The proposed methodology is based on multiple linear regression modeling using, on the one hand, the official coffee yields of Dak Lak province and, on the other hand, phenological variables derived from the seasonal dynamics of the satellite-derived biophysical variables NDVI (normalized difference vegetation index), LAI (leaf area index), and FAPAR (fraction of absorbed photosynthetically active radiation).

*2.2. Phenological Variables from Remote Sensing Time Series*

2.2.1. Vegetation Biophysical Variables

Coffee yield forecasting is based on satellite imagery from Copernicus Hub 2022 (source: https://land.copernicus.vgt.vito.be (accessed on 22 December 2021)), namely NDVI, LAI, and FAPAR, available at a decadal (10-day) time step for the entire study area in Dak Lak province and the same years as the official coffee yield statistics (Table 1). The 2000–2020 time series of decadal LAI, NDVI, and FAPAR products (21 years × 36 dekads/year) derived from the SPOT-VEGETATION and PROBA-V instruments were used in this study. The products are freely available at a 1 km global spatial resolution.

Table 1. Remote sensing vegetation biophysical products used in this study and downloaded from Copernicus Global Land Service (CGLS) (https://land.copernicus.vgt.vito.be/ (accessed on 22 December 2021)).

Remote Sensing Vegetation Biophysical Products	Definition	Period
FAPAR	Fraction of absorbed photosynthetically active radiation	2000–2020
LAI	Leaf area index	2000–2020
NDVI	Normalized difference vegetation index	2000–2020
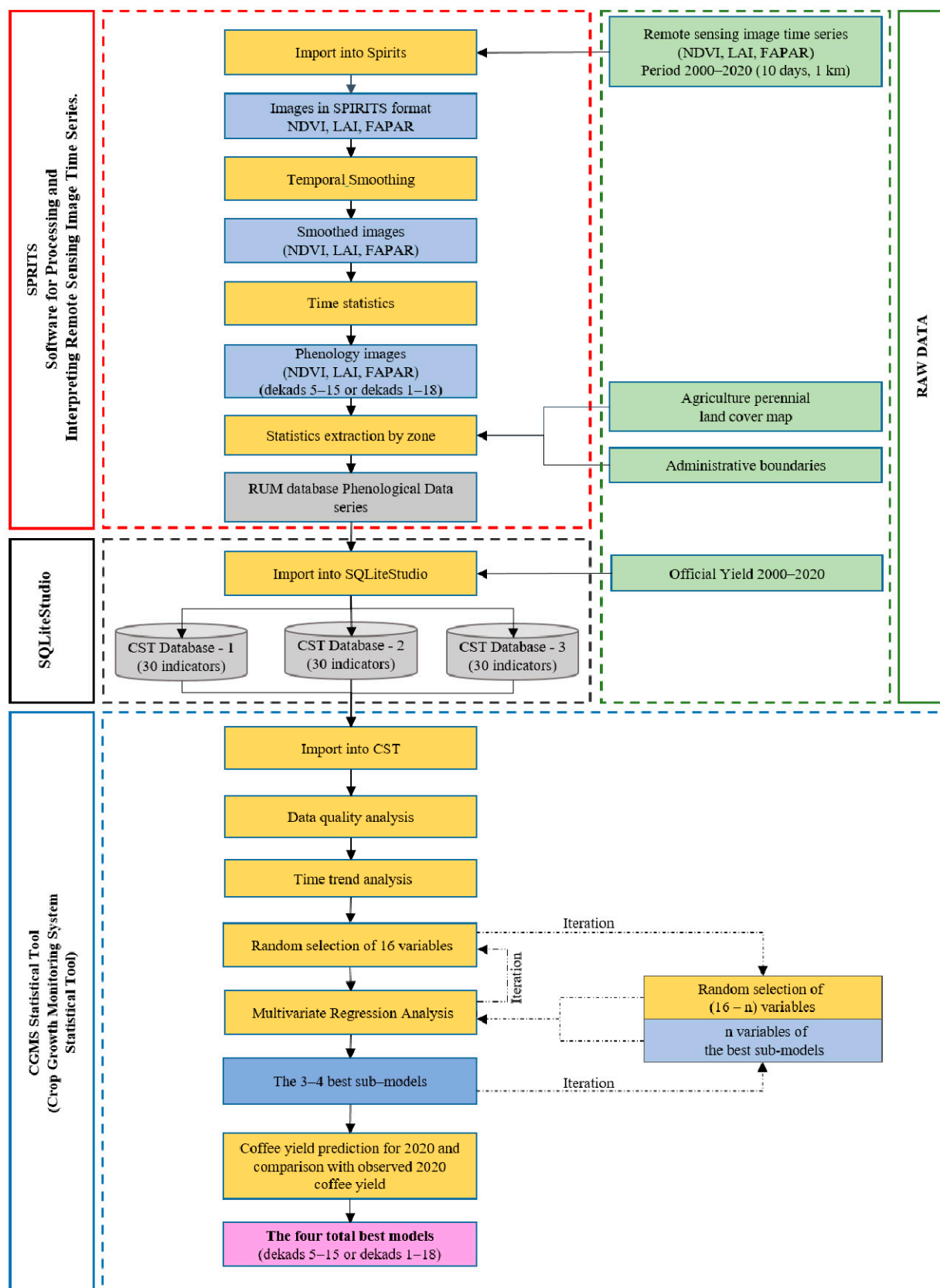
Figure 2. General workflow of the coffee yield forecasting method. Green rectangles: raw input data; yellow rectangles: data processing; blue rectangles: intermediate data; gray rectangles: variable databases; and pink rectangle: final results.

"The normalized difference vegetation index (NDVI) is an indicator of the greenness of the vegetation biomes". (source: https://land.copernicus.eu/global/products/ndvi (accessed on 22 December 2021))

NDVI has theoretical values ranging from −1 to +1, where negative values mostly correspond to clouds, water, and snow, whereas values near zero primarily correspond to rocks and bare soil [23]. NDVI increases progressively with vegetation development.

"The Leaf Area Index is defined as half the total area of green elements of the canopy per unit horizontal ground area. The satellite-derived value corresponds to the total green LAI of all the canopy layers, including the understory which may represent a very significant contribution, particularly for forests. Practically, the LAI quantifies the thickness of the vegetation cover." (source: https://land.copernicus.eu/global/products/lai (accessed on 22 December 2021))

"The FAPAR quantifies the fraction of the solar radiation absorbed by live leaves for the photosynthesis activity. Then, it refers only to the green and alive elements of the canopy." (source: https://land.copernicus.eu/global/products/FAPAR (accessed on 22 December 2021))

In Vietnam, coffee phenology can be divided into five periods: (i) the flower-bud initiation and blooming season is from January to March; (ii) the fruit-setting period is from April to May; (iii) the cherry development period is from May to August; (iv) the maturity stage is from September to October; and (v) the ripening/harvest period is from October to December [3]. Two periods were considered in this study to compute the explanatory variables used in the search for coffee yield prediction models. The first period corresponds to 11 dekads, from mid-February to the end of May (dekads 5 to 15). This period was considered because it corresponds to the crucial period of growth and development of the coffee bush [43]. February to May are normally dry months in Vietnam, and coffee requires irrigation to guarantee blossom and cherry settings [14]. The second period corresponds to 18 dekads, from January to June (dekads 1 to 18). This period was considered because a longer period may be more representative of the global coffee development conditions and consequently result in variables that have a higher explanatory power. Additionally, because the objective of the methodology developed in this research is to produce models that enable forecasting of coffee yield well in advance compared to the harvest period of October to December, we decided to set the coffee yield forecast as the end of June at the latest. Using the first six months of the year to predict coffee yield will give planners sufficient time to consider or find solutions before the end of the coffee season.

2.2.2. Processing of Satellite Images in SPIRITS Software

The NDVI, LAI, and FAPAR satellite image time series were processed in the free Software for Processing and Interpreting of Remote Sensing Image Time Series (SPIRITS) [44] (Figure 2).

First, images were imported and temporally smoothed with the SWETS algorithm [45], which was set with a maximum of 75% of missing values in each pixel profile, and the lowest physical value, Ymin, for cloud-free land pixels was kept at the default.

Second, the 11 phenological variables presented in Table 2 were computed from each of the 3 biophysical products (NDVI, LAI, and FAPAR), considering 2 periods (dekads 5 to 15 and dekads 1 to 18) by using the "time statistics" function of SPIRITS, which resulted in phenological images (Figure 2).

Third, zonal statistics were extracted for these phenological variable images for the perennial agricultural vegetation zone of Dak Lak province thanks to an extraction mask derived from the official 2015 land use map of Dak Lak province collected by the Department of Agriculture and Rural Development of Dak Lak province (Figure 1). This land use map did not contain a class specific to coffee plants but only a broad class relative to agricultural perennial plants accounting for approximately 62.5 to 68.2% of coffee from 2015 to 2018 [40,46–48]. No pure coffee crop mask was available, and it was not possible

for the authors to produce such a mask within the framework if this study. The extracted statistics corresponded to the final 33 coffee yield predictors.

Table 2. The 11 phenological variables derived from FAPAR, LAI, and NDVI time series (extracted using the time statistics function of SPIRITS [44] for 2 periods: dekads 5 to 15 and dekads 1 to 18, from 2000 to 2020).

No.	Variable	Definition	Dekads
1	vav	Average value (or mean)	5–15; 1–18
2	vmn	Minimum value	5–15; 1–18
3	vmx	Maximum value	5–15; 1–18
4	aup	Largest increase between subsequent periods	5–15; 1–18
5	adn	Largest decrease between subsequent periods	5–15; 1–18
6	rsd	Relative standard deviation (with N as denominator, not N − 1)	5–15; 1–18
7	rrg	Relative range (maximum–minimum)	5–15; 1–18
8	dmn	Relative date of (first) minimum value	5–15; 1–18
9	dmx	Relative date of (last) maximum value	5–15; 1–18
10	dup	Relative date of (first) largest increase	5–15; 1–18
11	ddn	Relative date of (last) largest decrease	5–15; 1–18

2.3. Official Coffee Yield Datasets

The coffee yields considered in this study were provincial coffee yields and were computed by dividing the official provincial coffee production by the official provincial coffee area according to the Dak Lak Statistical Yearbook (2009, 2014, 2018, and 2020) [40,46–48]. The period from 2000 to 2020 was considered. These coffee yields correspond to coffee dry grain yield.

2.4. Crop Yield Forecasting Model in the CST Software

In this study, the free software "Crop Growth Monitoring System Statistical Tool" (CgmsStatTool—CGMS statistical tool—CST) [28] was used to generate the coffee yield forecasting models.

The CST approach that we used in this study was multivariate regression analysis in order to assess the linear relationship between coffee yield (Y) and one or more independent variable(s) (the predictor(s) X_1, X_2 ...) with the following Equation (1):

$$Y \; = \; \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon \tag{1}$$

In Equation (1), ε is the random error assumed to follow a normal distribution of mean 0 and constant variance σ_2. Errors for different years are assumed to be independent. In the annotation for the X variables, the subscript n represents which X variable it is. $\beta_0 \ldots \beta_n$ are the regression coefficients to be calculated through the ordinary least square method, minimizing the difference between the observed and fitted yield values. The CST tests various models, potentially using the crop yield time trend and between 1 and 4 independent variables; then, standard statistics and plots are exported to enable assessment of the quality of these models.

CST analysis was carried out as follows: (1) check for possible errors in the database of official yields and indicators, (2) assess both linear and quadratic crop yield time trends at a significance level of 0.025, (3) assess the correlation between the indicators with and without time trend (if any), and (4) search for the best multivariate regression models.

> "CST takes the potential time trend into account by adding a term in the model that corresponds to that time trend, if applicable. To increase numerical precision, the regression coefficient for the linear time trend is for "year-offset" rather than "year" itself. The offset is fixed at 1965 by default in CST. Likewise, the regression coefficient for the quadratic time trend is for (year-offset)^2." [28]

The period of 2000–2019 was used to search for and build the best models through the multivariate regression method of the CST.

As the CST can only work with a database containing a maximum of 30 variables, 3 databases of 30 variables were built from the 33 input variables and were used sequentially. With 20 years of calibration data (2000–2019), the CST allows 16 variables to be tested at a time in the regression analysis. Therefore, we iterated a random selection of 16 input variables to find the best models (Figure 2).

The automatic selection and ordering of the best models by CST at each CST iteration for a given set of candidate variables was based on the root mean square error of prediction (RMSEp) (Equation (2)). RMSEp indicates the model's quality under prediction conditions [23]. RMSEp calculated by the CST is based on the leave-one-out residual or PRESS residual [28]. Predictions become increasingly precise as RMSEp approaches 0 and R^2 approaches 1. With each CST iteration, the 3 to 4 best sub-models were manually selected based on the RMSEp and the adjusted coefficient of determination (Adj.R^2) (Equation (3)). Adj.R^2 is a statistical measure of the model's goodness of fit in a regression model, which shows the proportion of variation explained by the estimated regression line.

$$\text{RMSEp} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} * \tag{2}$$

where:

P_i and O_i are the predicted and observed values for year i, respectively;

$\overline{O}$ and $\overline{P}$ express the means of observed and predicted values, respectively;

n is the number of samples (years); and

k is the number of independent variables in the regression equation.

* Note: in Equation (2), $P_i - O_i$ is the difference between the ith observation and the predicted value for the ith observation based on a model fit to the remaining observations, i.e., without the ith observations (adapted from [28]).

$$\text{Adj.}R^2 = 1 - \left(1 - R^2\right)\left[\frac{n-1}{n-(k+1)}\right] \tag{3}$$

Four other statistical parameters (Equations (4)–(7)) were also used to appreciate the models' performance but not to select them.

R squared (R^2) corresponds to the percentage of variance explained by the model (Equation (4)) [23].

$$R^2 = \left(\frac{\sum\limits_{i=1}^{n}(O_i - \overline{O})(P_i - \overline{P})}{\sqrt{\sum\limits_{i=1}^{n}(O_i - \overline{O})^2}\sqrt{\sum\limits_{i=1}^{n}(P_i - \overline{P})^2}}\right)^2 \tag{4}$$

The relative root mean square error (RRMSE) is calculated by dividing RMSEp by the mean value of observed data (Equation (5)).

$$\text{RRMSE}(\%) = \frac{\text{RMSEp}}{\overline{O}} \times 100 \tag{5}$$

The mean absolute percentage error (MAPE) is expressed as follows (Equation (6)).

$$\text{MAPE} = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(\frac{|O_i - P_i|}{|O_i|}\right) \times 100 \tag{6}$$

The residual standard deviation (RSD) is the square root of the residual mean square [28] (Equation (7)).

$$RSD = \sqrt{\dfrac{\sum\limits_{i=1}^{n} (P_i - O_i)^2}{df}} \qquad (7)$$

where df is degrees of freedom. Here, df is equal to the sample size minus the number of parameters in the model. For example: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$; therefore, df = n − 3, where n is the sample size, and the number of parameters is 3.

The year 2020 was used to assess the performance of selected models with an independent year not used in model calibration by comparing the observed and predicted yield for 2020 and computing the related residuals.

The final selection of the best models was based on a combination of model performance in calibration (2000–2019) and in prediction for 2020.

3. Results

3.1. Model Performance

According to the CST time trend analysis mode, Dak Lak province showed a significant upward linear time trend (*p*-value of 0.0012) for coffee yields during the period from 2000 to 2019 (Figure 3).
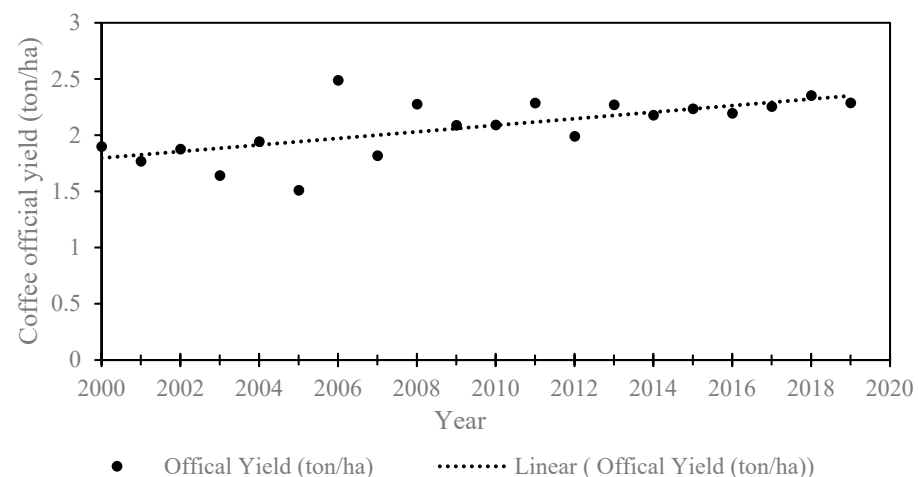
Figure 3. Time trend for official coffee yield of Dak Lak province for the period from 2000 to 2019.

Details of the eight best coffee yield models for Dak Lak province obtained with the multivariate regression method of the CST for the two periods considered (mid-February to end of May and early January to end of June) are presented in Table 3 and Figure 4.

Overall, the forecast coffee yield models performed satisfactorily in both time periods, with the RMSEp varying between 0.155 and 0.178 ton/ha, the RRMSE varying between 7.5% and 8.6%, and the Adjusted-R^2 varying between 62.8% and 68.8% (Table 3 and Figure 4). The models built on the 18-dekad period provided systematically better results than those built on the 11-dekad period when considering RMSEp and RRMSE only. For models computed based on 18 dekads, the RMSEp ranged from 0.155 to 0.158 ton/ha, the Adj.R^2 was between 64.2 and 68.8%, and the RRMSE ranged from 7.5 to 7.6. For models calculated based on 11 dekads, the RMSEp ranged from 0.174 to 0.178 ton/ha, the Adj.R^2 was between 62.8 and 67.6%, and the RRMSE ranged from 8.4 to 8.6%.

Table 3. Details of the eight best coffee yield models for Dak Lak province based on phenological variables derived from NDVI, FAPAR, and LAI for the 2000–2019 time period, with their related statistical performances. Models 1 to 4 are based on dekads 1 to 18 (January to June), and models 5 to 8 are based on dekads 5 to 15 (mid-March to end of May). Model 0 corresponds to the model achieved with the coffee yield linear time trend only.

	Parameter	Estimate	s.e.	t Value	RMSEp (ton/ha)	RRMSE (%)	R^2 (%)	Adj.R^2 (%)	MAPE (%)	RSD (ton/ha)
Model 0 **No dekad**	Constant	0.774	0.342	2.26	0.202	9.7		41.8	4.3	0.197
	Time trend (linear)	0.029	7.63×10^{-3}	3.83			44.8			
Model 1 **Dekads 1–18**	Constant	0.496	0.361	1.37	0.155	7.5		64.2	4.3	0.154
	Time trend (linear)	0.018	7.30×10^{-3}	2.46						
	dmx-LAI	−0.013	3.84×10^{-3}	−3.48			71.7			
	rrg-LAI	0.048	0.016	2.91						
	vmn-LAI	0.727	0.231	3.14						
Model 2 **Dekads 1–18**	Constant	0.494	0.362	1.37	0.155	7.5		64.5	4.3	0.154
	Time trend (linear)	0.018	7.30×10^{-3}	2.47						
	dmx-LAI	−0.013	3.85×10^{-3}	−3.47			71.7			
	vmn-LAI	0.155	0.186	0.83						
	vmx-LAI	0.572	0.197	2.91						
Model 3 **Dekads 1–18**	Constant	0.415	0.346	1.2	0.155	7.5		64.9	3.9	0.153
	Time trend (linear)	0.019	7.39×10^{-3}	2.56						
	dmx-LAI	-8.74×10^{-3}	4.27×10^{-3}	−2.05			72.3			
	dmx-NDVI	-3.58×10^{-3}	3.53×10^{-3}	−1.01						
	vmx-LAI	0.567	0.194	2.93						
Model 4 **Dekads 1–18**	Constant	1.708	0.592	2.88	0.158	7.6	75.4	68.8	4.7	0.144
	Time trend linear	0.013	6.79×10^{-3}	1.92						
	ddn-LAI	0.011	3.00×10^{-3}	3.65						
	rsd-FAPAR	−0.119	0.046	−2.56						
	vmn-LAI	0.35	0.139	2.52						
Model 5 **Dekads 5–15**	Constant	−0.344	0.412	−0.84	0.174	8.4		62.8	4.7	0.157
	Time trend (linear)	0.015	7.86×10^{-3}	1.85						
	adn-LAI	0.152	0.062	2.43			72.6			
	ddn-NDVI	9.96×10^{-3}	4.54×10^{-3}	2.19						
	dmn-LAI	0.019	6.24×10^{-3}	2.97						
	vmx-LAI	0.382	0.144	2.66						
Model 6 **Dekads 5–15**	Constant	3.218	0.972	3.31	0.177	8.5	76.1	67.6	5.6	0.147
	Time trend (linear)	0.025	6.10×10^{-3}	4.13						
	adn-NDVI	−9.122	2.602	−3.51						
	ddn-LAI	−0.01	4.03×10^{-3}	−2.55						
	ddn-NDVI	0.014	4.44×10^{-3}	3.05						
	dmx-FAPAR	−0.028	9.69×10^{-3}	−2.9						
Model 7 **Dekads 5–15**	Constant	1.917	0.738	2.6	0.178	8.6	73.2	63.6	5.0	0.156
	Time trend (linear)	0.015	9.21×10^{-3}	1.67						
	adn-LAI	0.081	0.039	2.06						
	aup-FAPAR	−1.74	0.673	−2.58						
	rrg-NDVI	−0.067	0.043	−1.55						
	rsd-LAI	0.152	0.055	2.79						
Model 8 **Dekads 5–15**	Constant	1.878	0.754	2.49	0.178	8.6	72.7	62.9	5.3	0.157
	Time trend (linear)	0.016	9.36×10^{-3}	1.69						
	adn-LAI	0.081	0.04	2.04						
	aup-FAPAR	−1.806	0.67	−2.7						
	rsd-LAI	0.165	0.063	2.6						
	rsd-NDVI	−0.221	0.151	−1.46						

s.e. = standard error, Adj.R^2 = adjusted R squared, R^2 = R squared, RSD = residual standard deviation, RMSEp = root mean square error for prediction, RRMSE = relative root mean square error (%).

It seems difficult to clearly identify one best model among those of the 18-dekad period, given that they all achieved very similar global statistical performance when considering all statistical parameters. For example, for the 18-dekad period, the best model according to the Adj.R^2 (model 4, Adj.R^2 of 68.8%) is the worst according to the RMSEp (0.158 ton/ha).
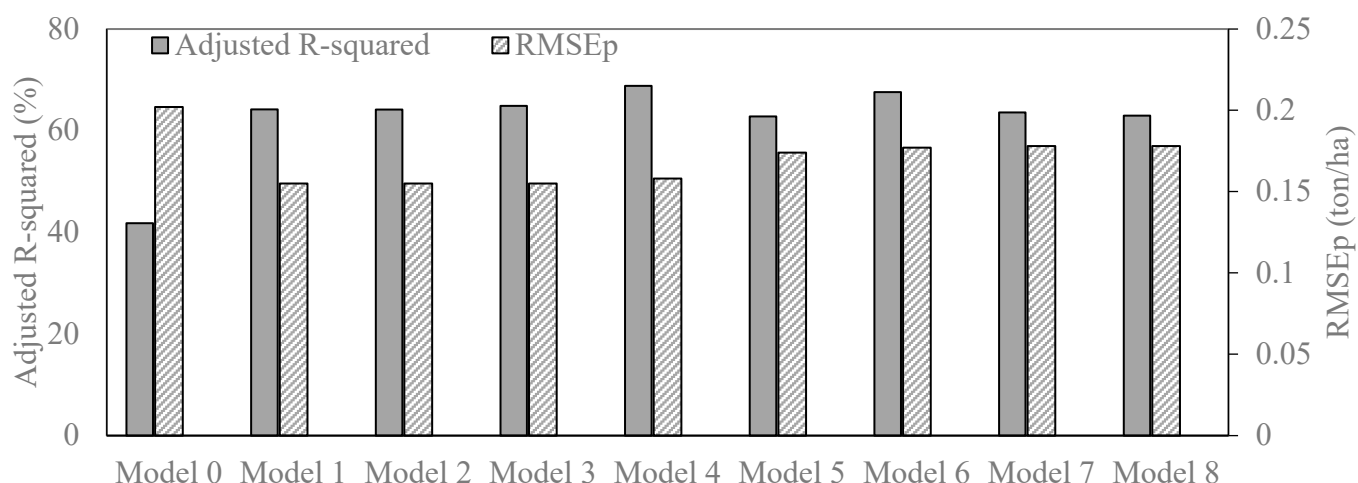
Figure 4. Adjusted R squared and RMSEp of the eight best coffee yield models based on phenological variables derived from NDVI, FAPAR, and LAI for the 2000–2019 time period. Models 1 to 4 are based on the dekads 1 to 18 (January to June), and models 5 to 8 are based on dekads 5 to 15 (mid-March to end of May). Model 0 corresponds to the model achieved with the coffee yield linear time trend only.
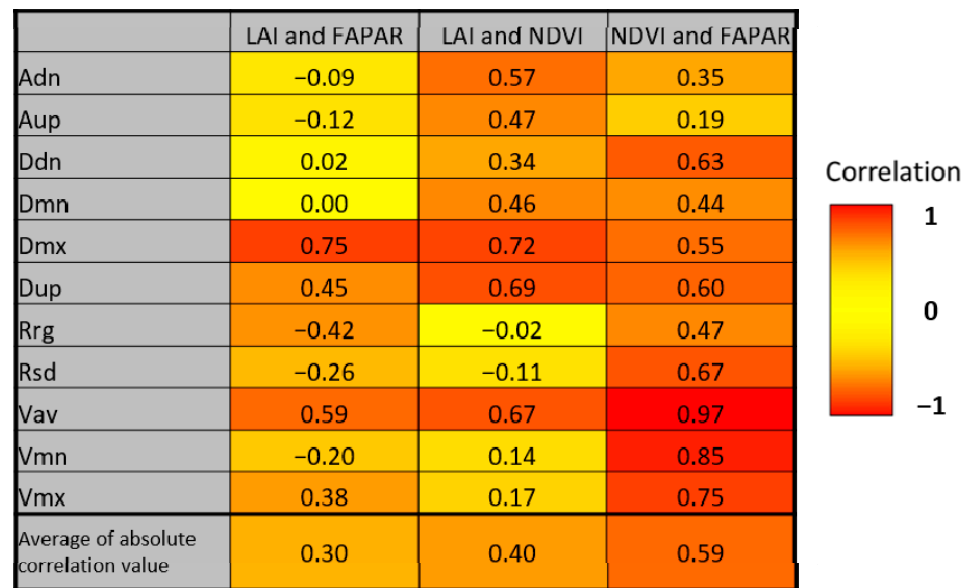
The results show that model 0, which corresponded to the linear time trend only, performed less efficiently (RMSEp = 0.202 ton/ha, RRMSE = 9.7%, Adj.R^2 = 41.8%) than models combining a linear time trend with phenological variables derived from the remote sensing data (Table 3 and Figure 4).

For the period considering dekads 1–18 (from January to June), all selected models used three variables, in addition to the time trend. Models 1 and 2 used only the LAI variables, model 3 combined the LAI and NDVI variables, and model 4 combined the LAI and FAPAR variables. For the period considering dekads 5–15, all models used four explanatory variables, in addition to the time trend. Model 5 combined the LAI and NDVI variables, whereas models 6–8 combined the LAI, NDVI, and FAPAR variables. Among the 8 best models, LAI-derived variables occur 18 times, whereas NDVI-derived variables occur 6 times and FAPAR-derived variables occur only 4 times. This observation suggests that LAI-derived variables are more efficient than NDVI- and FAPAR- derived variables for coffee yield forecasting. A relatively high negative or positive correlation was observed between some variables selected in some of the best models (R varies in the range from −0.863 (for Adn_LAI and Dmn_LAI in model 5) to 0.795 (for Vmn_LAI and Dmx_LAI in model 1)) Figure 5. When considering the period of dekad 1 (start of January) to dekad 18 (end of June) of the years 2000 to 2019, the analysis of the Pearson correlation coefficient of the 11 phenological variables between the three biophysical satellite products, LAI, NDVI, and FAPAR (Figure 6), shows a highly variable level of correlation between these phenological variables, from, in absolute values, 0.00 to 0.97, i.e., from no correlation to a very high level of correlation. For this period, the phenological variables derived from FAPAR and NDVI are the most correlated (average absolute correlation of 0.59; third column of Figure 6), whereas those derived from LAI are less correlated to NDVI and FAPAR variables, especially for FAPAR (average absolute correlation of 0.30; first column of Figure 6). The low correlation values observed for at least some phenological variables in each pair of biophysical products (LAI and FAPAR, LAI and NDVI, and FAPAR and NDVI) suggest that these three products may provide some non-redundant (uncorrelated) information and thus be complementary at some point and, consequently, that it is relevant to consider the three of them in the search for the best coffee yield prediction models. The results also showed that models utilizing satellite data from January to June (models 1 to 4) were more suitable for estimating coffee yields in Dak Lak province than models using satellite data from mid-February to May (lower RMSEp and higher Adj-R^2 for models 1 to 4).
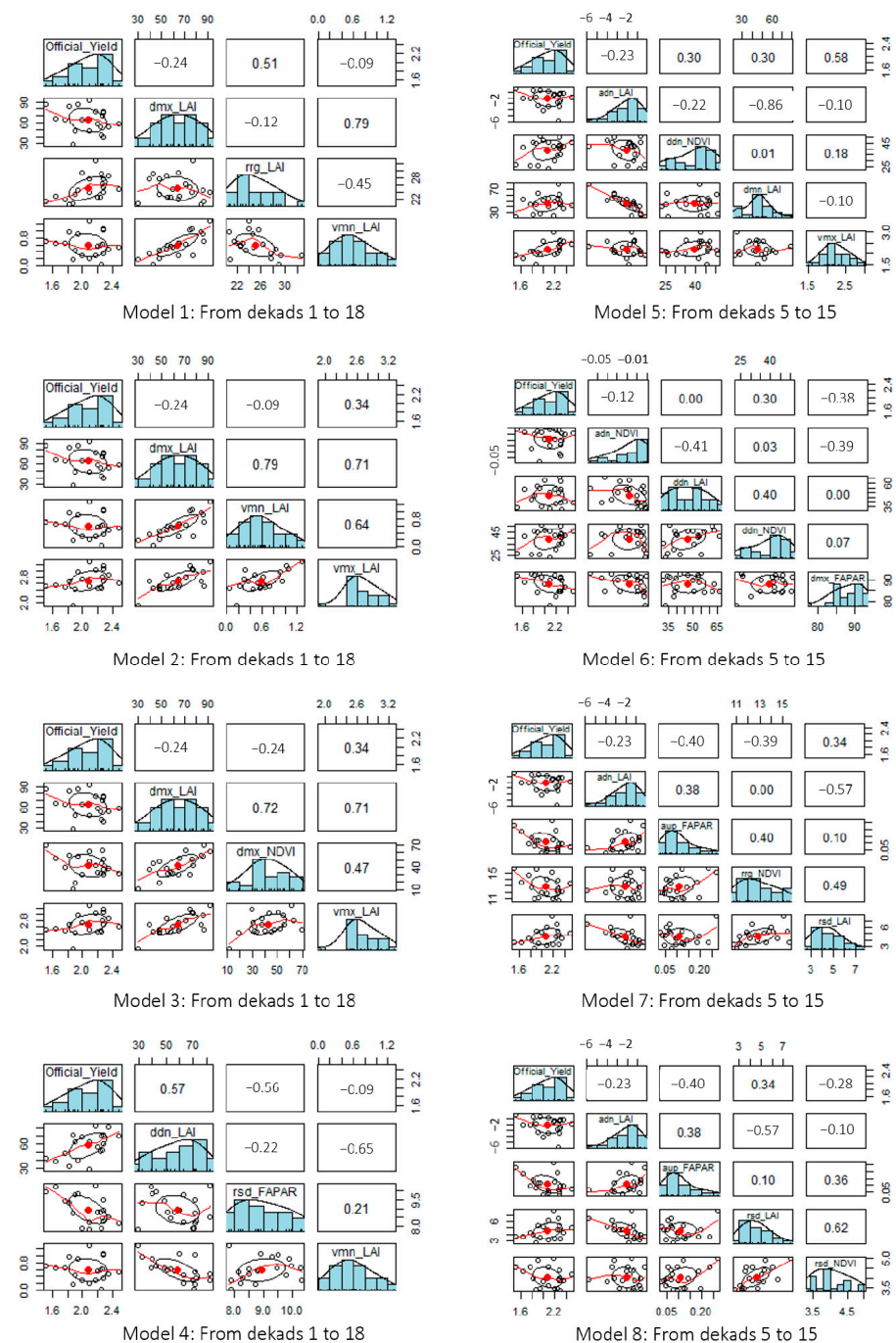
Figure 5. Correlation between the phenological variables derived from NDVI, FAPAR, and LAI for the 2000–2019 time period and selected in the eight best coffee yield models. Models 1 to 4 are based on dekads 1 to 18 (January to June), and models 5 to 8 are based on dekads 5 to 15 (mid-March to end of May). The values in the upper-right parts of the plots (above the diagonal) are the Pearson correlation coefficients between the two variables intersecting the corresponding row and column. On the diagonal is the histogram of each variable, which shows the lowest locally fit regression line. The plots below the diagonal are the bivariate scatter plots of each pair of variables. These scatter plots show an ellipse around the mean (the red point), with the axis length reflecting one standard deviation of the column and row variables. The red line is the smoothed regression lines of the bivariate scatter plots of each pair of variables.

	LAI and FAPAR	LAI and NDVI	NDVI and FAPAR
Adn	−0.09	0.57	0.35
Aup	−0.12	0.47	0.19
Ddn	0.02	0.34	0.63
Dmn	0.00	0.46	0.44
Dmx	0.75	0.72	0.55
Dup	0.45	0.69	0.60
Rrg	−0.42	−0.02	0.47
Rsd	−0.26	−0.11	0.67
Vav	0.59	0.67	0.97
Vmn	−0.20	0.14	0.85
Vmx	0.38	0.17	0.75
Average of absolute correlation value	0.30	0.40	0.59

Figure 6. Pearson correlation coefficient of the 11 phenological variables computed over the period of dekad 1 (start of January) to dekad 18 (end of June) of the years 2000 to 2019 between the three biophysical satellite products LAI, NDVI, and FAPAR.

3.2. Coffee Yield Predictions for 2020

Table 4 shows the residuals and percentage residuals of predicted coffee yields for the target year 2020 for the eight selected models. For models based on dekads 1 to 18 (models 1 to 4), the absolute residuals were in the range of 0.054 to 0.134 ton/ha, and the absolute percentage residuals were in the range of 2.2 to 5.5%. For models based on dekads 5 to 15, three models presented absolute residuals in the range of 0.248 to 0.571 ton/ha and absolute percentage residuals in the range of 10.2 to 23.6%, and one model (model 5) with a better performance presented a residual of 0.082 ton/ha and a percentage residual of 3.4%. The best model in terms of prediction for 2020 was model 3, with a residual of 0.054 ton/ha and a percentage residual of 2.2%. The 2020 residuals for models 1–4 (Table 4) were all smaller than the corresponding RMSEp of the period 2000–2019 (Table 3), whereas the 2020 residuals for models 5–8 (Table 4) were generally much higher than the corresponding RMSEp of the period 2000–2019 (Table 3).

Table 4. Coffee yield predictions for 2020 based on each model.

	Predicted Yield (ton/ha)	Official Yield (ton/ha)	Residual (ton/ha)	Percentage Residual (%)
Model 1	2.558	2.424	0.134	5.5
Model 2	2.558	2.424	0.134	5.5
Model 3	2.478	2.424	0.054	2.2
Model 4	2.298	2.424	−0.126	−5.2
Model 5	2.506	2.424	0.082	3.4
Model 6	2.995	2.424	0.571	23.6
Model 7	2.176	2.424	−0.248	−10.2
Model 8	2.171	2.424	−0.253	−10.4

Observed versus model-predicted coffee yields for the period 2000–2020 are presented in a series of scatterplots in Figure 7. The predicted values used in these plots are those predicted with the models calibrated for the 2000–2019 period.

These plots revealed the highest R^2 (0.76) for model 4, combining ddn-LAI, rsd-FAPAR, vmn-LAI, and yield linear time trend as predictor variables (Figure 7). For models based on data from January to June, the four selected models (model 1 to 4) indicated an R^2 in the range of 0.73 to 0.76 and a *p*-value of <0.0001 (Figure 7).
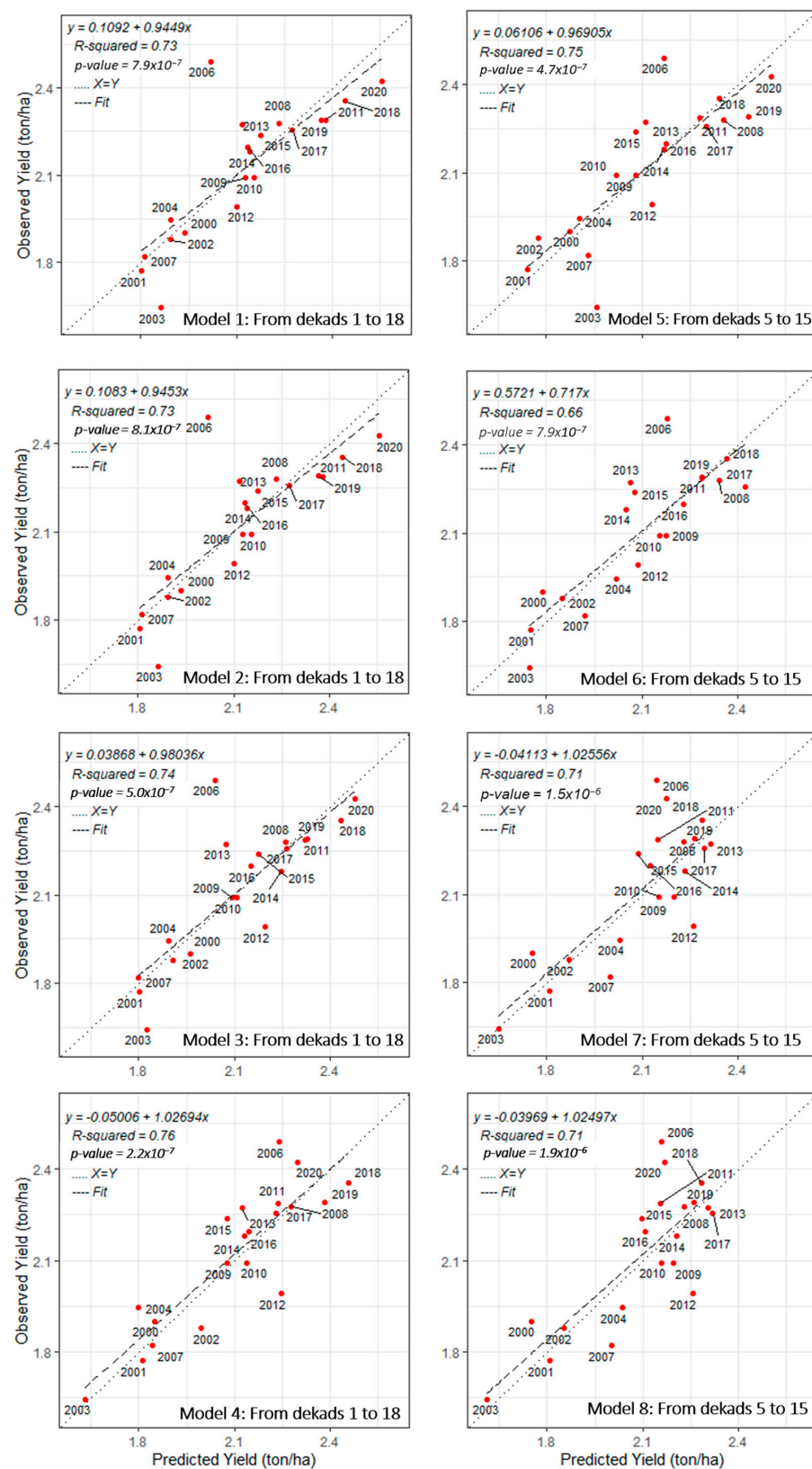
Figure 7. Scatter plot of observed versus model-predicted coffee yields for the years 2000 to 2020 using the eight best selected models based on satellite data for the period from dekads 1 to 18 (models 1 to 4) and for the period from dekads 5 to 15 (models 5 to 8). For model 6, the predicted value for the year 2020 is 3.0 ton/ha, which is outside the plot frame. R-squared and *p*-values reported in this figure are those of the relation between observed versus predicted yield for the full 2000–2020 period.

The models based on data from mid-February to May presented an R^2 ranging from 0.66 to 0.75 and a *p*-value of <0.0001 (Figure 7).

In 2006, most models underestimated the official yields by approximately 0.249 to 0.470 ton/ha.

4. Discussion

The observed positive coffee yield time trend in Dak Lak province over the past 20 years can be explained by a combination of factors, including investments in irrigation infrastructure, a heavy reliance on irrigation on coffee farms, the affordability of fertilizer, and the increasing adoption of new management techniques in the province [3,42].

Existing coffee models usually simulate and forecast coffee yields at a local or regional level by including, as parameters, the main growth and development processes impacted by climate variations. However, in this study, the results showed that it is possible to predict coffee yield at a regional scale in Dak Lak province of Vietnam six months before the harvest based on remote sensing data only. Thus, such models could be an essential tool for indirectly assessing the impacts of weather variability or improvements in farmer practices on coffee yields at the provincial or district level in Vietnam or any other coffee-growing regions or countries under climate change conditions.

Figures 3 and 5 show that 2006 presented with the highest observed coffee yield, which also corresponded to the year with the lowest model accuracy. We have no information that could explain such a high yield in 2006. In particular, the precipitation in 2006 was not particularly high.

Compared to the coffee yield forecast models developed by Kouadio et al. (2018) [21] and Kouadio et al. (2021) [3], most of the models developed in this study achieved a higher accuracy (RMSEp = 0.155 to 0.178 ton/ha, RMSE = 0.123 to 0.134 ton/ha, RRMSE = 7.5 to 8.6%, and MAPE = 3.9 to 5.3%) (Table 3). The model of Kouadio et al. (2018), which consisted of an extreme learning machine (ELM) model using soil organic matter (SOM), available potassium, and available sulfur as explanatory variables, provided coffee yield estimates for Lam Dong province (in the central highlands of Vietnam) with an RMSE of 0.496 ton/ha, an RRMSE of 13.6%, and MAPE = 7.9% [21]. In addition, the simple process-based model developed by Kouadio et al. (2021) for simulating and forecasting robusta coffee yield at the regional scale in Vietnam showed a RMSE of 0.24 to 0.33 ton/ha and an MAPE = 9 to 14%, and that model was successfully tested using satellite remote sensing data (LAI) and model-based gridded climate data (maximum and minimum temperatures, solar radiation, and rainfall): MAPE $\leq$ 12% and RMSE $\leq$ 0.29 ton/ha [3].

The method proposed in this study enabled the development of models that can satisfactorily forecast coffee yield with a low RMSEp and a high Adj.R^2 for Dak Lak province. We also showed that the period considered for the production of the model explanatory variables (dekads 1 to 18 versus dekads 5 to 15) has an important impact on the accuracy of the resulting models, with better accuracy for those considering a more extended period. The models based on a longer period were also composed of fewer explanatory variables (three variables + the time trend) than those based on a shorter period (four variables + the time trend).

When selected in models, the variables aup-FAPAR, dmx-LAI, dmx-FAPAR, and dmx-NDVI presented systematically negative values, which may mean that the smaller the "largest increase between subsequent periods" of the FAPAR and the sooner the "date of maximum" LAI, FAPAR, and NDVI, the higher the coffee yield will be. Furthermore, the variables derived from the LAI product were demonstrated to be more efficient for coffee yield forecast models than those derived from NDVI and FAPAR, although some complementarity was observed between these products for some models (Table 3).

This is the first research to date combining NDVI, FAPAR, and LAI remote-sensing-derived phenological variables in the CST to create a coffee yield prediction model, which is the main contribution of this study. The data used in this study were derived from the SPOT-VEGETATION and PROBA-V instruments at a 1 km global spatial resolution.

Therefore, future studies should consider using more recent and similar products derived at a 300 m spatial resolution.

The main technical limitations we encountered in this research are related to the fact that the CST cannot handle a database containing more than 30 variables and that with 20 years of data used for model calibration, the CST cannot consider more than 16 variables at a time during the multiple linear regression model search. These two technical limitations of the CST make its use more difficult than it should be.

In this research, we used a multiple linear regression technique in order to produce coffee yield prediction models. Such a technique is particularly suited to identify and use the linear relationships between the predictors and the dependent variable. However, the linearity of the relationship between coffee yield and the phenological variables was not assessed in this study, and it may be possible that some variables present a nonlinear relationship with yield. Consequently, further research should involve testing other nonlinear modelling approaches for predicting coffee yield from biophysical variables, such those used in this study.

The findings of this study show that satellite data, such as the NDVI, LAI, and FAPAR products provided by the Copernicus Global Land Service are a good source of information for estimating and forecasting coffee yield in a challenging situation, where there is a deficit of information about management practices, soil characteristics, irrigation schedule, phenology of coffee trees, etc.

We think that a main source of improvement of the coffee yield forecast model developed in this research would probably be the use of a more detailed land use map containing a class specific to coffee. The official 2015 land use map of Dak Lak province used to extract satellite-derived vegetation biophysical variables did not contain a class specific to coffee plants but only a broad class relative to agricultural perennial plants, accounting for approximately 62.5 to 68.2% of coffee from 2015 to 2018 [40,46–48].

5. Conclusions

The present study is the first research on the development and assessment of a coffee yield forecasting method at the regional scale for Dak Lak province in the central highlands of Vietnam using the Crop Growth Monitoring System Statistical Tool (CGMSstatTool—CST) software and vegetation biophysical variables (NDVI, LAI, and FAPAR) derived from satellite remote sensing data (SPOT-VEGETATION and PROBA-V).

The findings of this research reveal that the elaboration of multiple linear regression models based on satellite-derived vegetation biophysical variables (LAI, NDVI, and FAPAR) corresponding to the first six months of the years 2000–2019 resulted in coffee yield forecast models presenting satisfactory accuracy (Adj.R^2 = 64 to 69%, RMSEp = 0.155 to 0.158 ton/ha, and MAPE = 3.9 to 4.7%). These results demonstrate that the CST may efficiently predict coffee yields on a regional scale by using only satellite-derived vegetation biophysical variables. Our findings are likely to aid local governments and decision makers in precisely forecasting coffee production early and promptly, as well as in recommending relevant local agricultural policies.

Further research should consider applying the developed method to search for coffee yield forecast models at other scales (at district and national levels) with enhanced input data (finer spatial resolution for satellite images and more accurate coffee maps) and with other explanatory variables.

Author Contributions: Conceptualization, N.T.T.T., D.N.K., A.D. and B.T.; methodology, N.T.T.T., D.N.K., A.D. and B.T.; analysis, N.T.T.T.; writing—original draft, N.T.T.T.; writing—review and editing, N.T.T.T., A.D., D.N.K., L.V.V., J.W. and B.T.; supervision, D.N.K. and B.T.; project administration, D.N.K., L.V.V., J.W. and B.T. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The in situ datasets generated and/or analyzed during the current study and the satellite derived database are available from the corresponding author upon reasonable request. The geographical boundaries of the fields analyzed during the current study are not publicly available due to privacy protection. The satellite images are freely available at https://land.copernicus.vgt.vito.be/.

Conflicts of Interest: The authors declare no conflict of interest. The funder had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Ubilava, D. El Niño, La Niña, and World Coffee Price Dynamics. *Agric. Econ.* **2012**, *43*, 17–26. [CrossRef]
2. Pohlan, J.; Janssens, M. Growth and Production of Coffee. In *Soils, Plant Growth and Crop Production*; Verheye, W.H., Ed.; Encyclopedia of Life: Oxford, UK, 2015.
3. Kouadio, L.; Tixier, P.; Byrareddy, V.; Marcussen, T.; Mushtaq, S.; Rapidel, B.; Stone, R. Performance of a Process-Based Model for Predicting Robusta Coffee Yield at the Regional Scale in Vietnam. *Ecol. Model.* **2021**, *443*, 109469. [CrossRef]
4. DaMatta, F.M.; Rahn, E.; Läderach, P.; Ghini, R.; Ramalho, J.C. Why Could the Coffee Crop Endure Climate Change and Global Warming to a Greater Extent than Previously Estimated? *Clim. Chang.* **2019**, *152*, 167–178. [CrossRef]
5. Pham, Y.; Reardon-Smith, K.; Mushtaq, S.; Cockfield, G. The Impact of Climate Change and Variability on Coffee Production: A Systematic Review. *Clim. Chang.* **2019**, *156*, 609–630. [CrossRef]
6. Piato, K.; Lefort, F.; Subía, C.; Caicedo, C.; Calderón, D.; Pico, J.; Norgrove, L. Effects of Shade Trees on Robusta Coffee Growth, Yield and Quality. A Meta-Analysis. *Agron. Sustain. Dev.* **2020**, *40*, 38. [CrossRef]
7. USDA. Foreign Agricultural Service Volume of Coffee Exports from Vietnam from 2011 to 2021. Available online: https://www.statista.com/statistics/877329/vietnam-coffee-export-volume/ (accessed on 10 December 2021).
8. ICO. Coffee Production Worldwide in 2020, by Leading Country. Available online: https://www.statista.com/statistics/277137/world-coffee-production-by-leading-countries/ (accessed on 10 December 2021).
9. Thao, N.T.T.; Khoi, D.N.; Xuan, T.T.; Tychon, B. Assessment of Livelihood Vulnerability to Drought: A Case Study in Dak Nong Province, Vietnam. *Int. J. Disaster Risk Sci.* **2019**, *10*, 604–615. [CrossRef]
10. Jayakumar, M.; Rajavel, M.; Surendran, U.; Gopinath, G.; Ramamoorthy, K. Impact of Climate Variability on Coffee Yield in India—with a Micro-Level Case Study Using Long-Term Coffee Yield Data of Humid Tropical Kerala. *Clim. Chang.* **2017**, *145*, 335–349. [CrossRef]
11. Kath, J.; Byrareddy, V.M.; Craparo, A.; Nguyen-Huy, T.; Mushtaq, S.; Cao, L.; Bossolasco, L. Not so Robust: Robusta Coffee Production Is Highly Sensitive to Temperature. *Glob. Chang. Biol.* **2020**, *26*, 3677–3688. [CrossRef]
12. Nguyen, D.Q.; Renwick, J.; Mcgregor, J. Variations of Surface Temperature and Rainfall in Vietnam from 1971 to 2010. *Int. J. Climatol.* **2014**, *34*, 249–264. [CrossRef]
13. Van Viet, L. Development of a New ENSO Index to Assess the Effects of ENSO on Temperature over Southern Vietnam. *Theor. Appl. Climatol.* **2021**, *144*, 1119–1129. [CrossRef]
14. Vo, T. *Vietnam Coffee Annual 2020*; USDA: Washington, DC, USA, 2021; Volume 2021.
15. Gutierrez, A.P.; Villacorta, A.; Cure, J.R.; Ellis, C.K. Tritrophic Analysis of the Coffee (Coffea Arabica)—Coffee Berry Borer [Hypothenemus Hampei (Ferrari)]—Parasitoid System. *An. Soc. Entomol. Bras.* **1998**, *27*, 357–385. [CrossRef]
16. Rodríguez, D.; Cure, J.R.; Cotes, J.M.; Gutierrez, A.P.; Cantor, F. A Coffee Agroecosystem Model: I. Growth and Development of the Coffee Plant. *Ecol. Model.* **2011**, *222*, 3626–3639. [CrossRef]
17. Vezy, R.; le Maire, G.; Christina, M.; Georgiou, S.; Imbach, P.; Hidalgo, H.G.; Alfaro, E.J.; Blitz-Frayret, C.; Charbonnier, F.; Lehner, P.; et al. DynACof: A Process-Based Model to Study Growth, Yield and Ecosystem Services of Coffee Agroforestry Systems. *Environ. Model. Softw.* **2020**, *124*, 104609. [CrossRef]
18. Van Oijen, M.; Dauzat, J.; Harmand, J.M.; Lawson, G.; Vaast, P. Coffee Agroforestry Systems in Central America: II. Development of a Simple Process-Based Model and Preliminary Results. *Agrofor. Syst.* **2010**, *80*, 361–378. [CrossRef]
19. Rahn, E.; Vaast, P.; Läderach, P.; van Asten, P.; Jassogne, L.; Ghazoul, J. Exploring Adaptation Strategies of Coffee Production to Climate Change Using a Process-Based Model. *Ecol. Model.* **2018**, *371*, 76–89. [CrossRef]
20. Ovalle-Rivera, O.; Van Oijen, M.; Läderach, P.; Roupsard, O.; de Melo Virginio Filho, E.; Barrios, M.; Rapidel, B. Assessing the Accuracy and Robustness of a Process-Based Model for Coffee Agroforestry Systems in Central America. *Agrofor. Syst.* **2020**, *94*, 2033–2051. [CrossRef]

21. Kouadio, L.; Deo, R.C.; Byrareddy, V.; Adamowski, J.F.; Mushtaq, S.; Phuong Nguyen, V. Artificial Intelligence Approach for the Prediction of Robusta Coffee Yield Using Soil Fertility Properties. *Comput. Electron. Agric.* **2018**, *155*, 324–338. [CrossRef]

22. Molina, A.L.V.; Peralta, V.P.P.; Orozco, A.B.P.; Iglesias, M.I.O.; Guerrero, E.G. Calibration of the Aquacrop Model in Special Coffee (Coffea Arabica) Crops in the Sierra Nevada of Santa Marta, Colombia. *J. Agron.* **2018**, *17*, 241–250. [CrossRef]

23. Fall, C.M.N.; Lavaysse, C.; Kerdiles, H.; Dramé, M.S.; Roudier, P.; Gaye, A.T. Performance of Dry and Wet Spells Combined with Remote Sensing Indicators for Crop Yield Prediction in Senegal. *Clim. Risk Manag.* **2021**, *33*, 100331. [CrossRef]

24. Tebaldi, C.; Lobell, D.B. Towards Probabilistic Projections of Climate Change Impacts on Global Crop Yields. *Geophys. Res. Lett.* **2008**, *35*, 2–7. [CrossRef]

25. Laudien, R.; Schauberger, B.; Makowski, D.; Gornott, C. Robustly Forecasting Maize Yields in Tanzania Based on Climatic Predictors. *Sci. Rep.* **2020**, *10*, 19650. [CrossRef]

26. Nain, A.S.; Dadhwal, V.K.; Singh, T.P. Use of CERES-Wheat Model for Wheat Yield Forecast in Central Indo-Gangetic Plains of India. *J. Agric. Sci.* **2004**, *142*, 59–70. [CrossRef]

27. Lobell, D.B.; Burke, M.B. On the Use of Statistical Models to Predict Crop Yield Responses to Climate Change. *Agric. For. Meteorol.* **2010**, *150*, 1443–1452. [CrossRef]

28. Goedhart, P.W.; Hoek, S.B.; Boogaard, H.L. *The CGMS Statistical Tool*; Contributions by 2019; European Commission: Ispra, Italy, 2019.

29. Kerdiles, H.; Rembold, F.; Leo, O.; Boogaard, H.; Hoek, S. CST, a Freeware for Predicting Crop Yield from Remote Sensing or Crop Model Indicators: Illustration with RSA and Ethiopia. In Proceedings of the 2017 6th International Conference on Agro-Geoinformatics, Fairfax, VA, USA, 7–10 August 2017. [CrossRef]

30. Qing, H.; Fei, T.; Jianqiang, R.; Wenbin, W.; Dandan, L.; Hui, D. The Application of China-CGMS in the Main Crop Growth Monitoring in Northeast China. In Proceedings of the 2012 First International Conference on Agro- Geoinformatics (Agro-Geoinformatics), Shanghai, China, 2–4 August 2012; pp. 1–4.

31. Rembold, F.; Meroni, M.; Urbano, F.; Royer, A.; Atzberger, C.; Lemoine, G.; Eerens, H.; Haesen, D. Remote Sensing Time Series Analysis for Crop Monitoring with the SPIRITS Software: New Functionalities and Use Examples. *Front. Environ. Sci.* **2015**, *3*, 46. [CrossRef]

32. Balaghi, R.; Badjeck, M.C.; Bakari, D.; De Pauw, E.D.; De Wit, A.; Defourny, P.; Donato, S.; Gommes, R.; Jlibene, M.; Ravelo, A.C.; et al. Managing Climatic Risks for Enhanced Food Security: Key Information Capabilities. *Procedia Environ. Sci.* **2010**, *1*, 313–323. [CrossRef]

33. De Wit, A.; Duveiller, G.; Defourny, P. Estimating Regional Winter Wheat Yield with WOFOST through the Assimilation of Green Area Index Retrieved from MODIS Observations. *Agric. For. Meteorol.* **2012**, *164*, 39–52. [CrossRef]

34. Araya, S.; Ostendorf, B.; Lyle, G.; Lewis, M. Remote Sensing Derived Phenological Metrics to Assess the Spatio-Temporal Growth Variability in Cropping Fields. *Adv. Remote Sens.* **2017**, *06*, 212–228. [CrossRef]

35. López-Lozano, R.; Duveiller, G.; Seguini, L.; Meroni, M.; García-Condado, S.; Hooker, J.; Leo, O.; Baruth, B. Towards Regional Grain Yield Forecasting with 1km-Resolution EO Biophysical Products: Strengths and Limitations at Pan-European Level. *Agric. For. Meteorol.* **2015**, *206*, 12–32. [CrossRef]

36. 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*, 2492–2509. [CrossRef]

37. Nogueira, S.M.C.; Moreira, M.A.; Volpato, M.M.L. Relationship between coffee crop productivity and vegetation indexes derived from oli/landsat-8 sensor data with and without topographic correction. *Int. Braz. Assoc. Agric. Eng.* **2018**, *38*, 387–394. [CrossRef]

38. DakLak Provincial People's Committee. Available online: https://daklak.gov.vn/web/english/about-daklak (accessed on 4 April 2022).

39. CCAFS-SEA. *The Drought Crisis in the Central Highlands of Vietnam—Assessment Report*; CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS): Hanoi, Vietnam, 2016.

40. DakLak Statistical Office. *DakLak Statistical Yearbook 2020*; DakLak Statistical Office: DakLak, Vietnam, 2021.

41. Huong, N.T.; Anh, L.H. Factors Affecting the Technical Efficiency of Coffee Producers—Case Study in Dak Lak Province, Vietnam. *Int. J. Econ. Commer. Manag.* **2019**, *VII*, 535–543.

42. Byrareddy, V.; Kouadio, L.; Kath, J.; Mushtaq, S.; Rafiei, V.; Scobie, M.; Stone, R. Win-Win: Improved Irrigation Management Saves Water and Increases Yield for Robusta Coffee Farms in Vietnam. *Agric. Water Manag.* **2020**, *241*, 106350. [CrossRef]

43. Titus, A.; Pereira, G.N. Water Use Efficiency for Robusta Coffee. Available online: https://ecofriendlycoffee.org/water-use-efficiency-robusta-coffee/ (accessed on 17 December 2021).

44. Eerens, H.; Dominique, H. *Software for the Processing and Interpretation of Remotely Sensed Image Time Series*; User's Manual Version 1.5.2—February 2018; VITO, EU Joint Research Center: Ispra, Italy, 2018.

45. Swets, D.; Reed, B.C.; Rowland, J.; Marko, S.E. A Weighted Least-Squares Approach to Temporal NDVI Smoothing. In Proceedings of the From Image to Information: 1999 ASPRS Annual Conference, Portland, Oregon, 17–21 May 1999.

46. DakLak Statistical Office. *DakLak Statistical Yearbook 2009*; DakLak Statistical Office: DakLak, Vietnam, 2010.

47. DakLak Statistical Office. *DakLak Statistical Yearbook 2014*; DakLak Statistical Office: DakLak, Vietnam, 2015.

48. DakLak Statistical Office. *DakLak Statistical Yearbook 2018*; DakLak Statistical Office: DakLak, Vietnam, 2019.