sustainability

MDPI

Article

Prediction of Banana Production Using Epidemiological Parameters of Black Sigatoka: An Application with Random Forest

Barlin O. Olivares [1,*], Andrés Vega [2], María A. Rueda Calderón [3], Edilberto Montenegro-Gracia [4], Miguel Araya-Almán [5] and Edgloris Marys [6]

[1] Programa de Doctorado en Ingeniería Agraria, Alimentaria, Forestal y del Desarrollo Rural Sostenible, Universidad de Córdoba, Carretera Nacional IV, km 396, 14014 Córdoba, Spain
[2] Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba, Av. Haya de la Torre s/n, Córdoba X5000HUA, Argentina
[3] Laboratorio de Genética y Genómica Aplicada, Escuela de Ciencias del Mar, Pontificia Universidad Católica de Valparaíso, Chile. S/Brasil, Valparaíso 2950, Chile
[4] Facultad de Ciencias Agropecuarias, Universidad de Panamá, CRUBO Bocas del Toro, Finca 15, Changuinola 01001, Panama
[5] Departamento de Ciencias Agrarias, Universidad Católica del Maule, km 6 Camino Los Niches, Curicó 3466706, Chile
[6] Laboratorio de Biotecnología y Virología Vegetal, Centro de Microbiología y Biología Celular, Instituto Venezolano de Investigaciones Científicas (IVIC), Caracas 1204, Venezuela
* Correspondence: barlinolivares@gmail.com

Abstract: Accurate predictions of crop production are critical to developing effective strategies at the farm level. Knowing banana production is due to the need to maximize the investment–profit ratio, and the availability of this information in advance allows decisions to be made about the management of important diseases. The objective of this study was to predict the number of banana bunches from epidemiological parameters of Black Sigatoka (BS), using random forests (RF) for its ability to predict crop production responses to epidemiological variables. Weekly production data (number of banana bunches) and epidemiological parameters of BS from three adjacent banana sites in Panama during 2015–2018 were used. RF was found to be very capable of predicting the number of banana bunches, with variance explained as 70.0% and root mean square error (RMSE) of 1107.93 ± 22 of the mean banana bunches observed in the test case. The site, week, youngest leaf spotted and youngest leaf with symptoms in plants with 10 weeks of physiological age were found to be the best predictor group. Our results show that RF is an efficient and versatile machine learning method for banana production predictions based on epidemiological parameters of BS due to its high accuracy and precision, ease of use, and usefulness in data analysis.

Keywords: Black Sigatoka; Musa; production; banana disease; random forest; machine learning

check for updates

Citation: Olivares, B.O.; Vega, A.; Rueda Calderón, M.A.; Montenegro-Gracia, E.; Araya-Almán, M.; Marys, E. Prediction of Banana Production Using Epidemiological Parameters of Black Sigatoka: An Application with Random Forest. *Sustainability* **2022**, *14*, 14123. https://doi.org/10.3390/su142114123

Academic Editor: Mariolina Gullì

Received: 2 October 2022
Accepted: 27 October 2022
Published: 29 October 2022

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

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

1. Introduction

The banana (*Musa* spp.) economically and globally represents the most important fruit species in production and trade [1,2]. In the case of Panama, it is one of the largest producers in Latin America, ranking tenth in the FAO ranking [1,3] with a production of 308,835 tons, a cultivated area of 8449 ha, and a yield of 36,553 Kg/ha, which makes it an essential crop both for its economic importance and for its deep cultural roots [4,5].

However, banana production has been seriously harmed by the emergence of Black Sigatoka (BS), or black streak disease, of banana leaves, caused by the pathogenic fungus *Pseudocercospora fijiensis* (asexual phase, [Morelet] Deighton; the sexual phase of *P. fijiensis* is *Mycosphaerella fijiensis* Morelet) [6–8]. The control of this disease in Central and South America represents 27% of production costs, this being approximately USD 10 million

annually. It is for this reason that BS is considered the most severe and destructive banana disease in the world [9,10], causing the largest harvest losses in banana plantations and outweighing all other banana fungal diseases and the high costs for their control [11–13].

These financial limitations, coupled with the decrease in production in many local farmers in developing countries, are the basis for generating studies that seek to predict banana production based on BS epidemiological parameters [7,14]. Therefore, predicting the next harvests through rates of BS infections in banana plants would allow farmers to take appropriate preventive measures and mitigate disease management costs. Although there are studies linking BS forecasts to meteorological variables [15–17], powerful models to forecast BS infections have not yet been developed (so far as we know), much less to predict banana production.

Advances in machine learning algorithms in bananas, such as neural networks (NN) [18–20], random forest (RF) [11,20–23], support vector machines (SVM) [21,24,25], and orthogonal partial least-squares-discriminant analysis (OPLS-DA) [11], have addressed time series models. However, these models have difficulty modeling irregularity over time. Recently, a study by Olivares et al. [11] established through supervised methods such as RF and OPLS-DA relationships between incidences of banana wilt and soil properties, demonstrating the great classifying and predictive power that RF has in this type of situation. Today, RF regression applications in crop science remain lacking, with few exceptions. Numerous studies have pointed out various promising advantages of RF as a regression tool compared to traditional regression models [21–23]; therefore the initiative to use the RF algorithm arises in this study focused on its usefulness as a prediction tool in banana production.

For the prediction of yield (number of bunches), the farmer's experience is important when there are no methods for estimation and this becomes the only resource; however, these approximations may be insufficient, and what is needed is systematically stored information that an example includes, first of all, average historical production records; and, secondly, yield variations due to agricultural management, climatic factors, epidemiological parameters or disease management, among others, to reduce any bias or error. This situation causes new methodologies to be developed and other data to be considered, for example: physical, environmental, epidemiological, or climatic, that can improve the quality of predictions.

Although various scientific studies have produced mathematical models of infectious diseases, it is difficult to adapt these models to incorporate the production approach in terms of the number of banana bunches or the number of exportable boxes, for example. This study aimed to validate the hypothesis that BS epidemiological parameters can predict banana bunch numbers in large areas of Bocas del Toro, Panama using an RF prediction model. In this case, the quantitatively estimated parameters of this disease could be used to improve the evaluation of banana production. Analysis used with highly predictive RF can provide a working model for bananas in an area where there is little history of such problems.

2. Materials and Methods

2.1. Study Area

The study refers to three commercial banana farms belonging to the same cooperative, with banana plants (AAA) Cavendish cv. Williams and Gran Enano, located in the Changuinola District of the Bocas del Toro province in Panama (Table 1). The climate is tropical and rainy, and the amount of precipitation fluctuates between 2500 mm annually and has a non-seasonal rainfall regime, where the lowest rainfall is in February (143 mm), and most of the precipitation falls in December (284 mm). The absolute maximum temperature is 36 °C and the minimum is 15 °C, with an annual mean temperature between 25–26 °C [26].

Table 1. The geographical location of cultivated area (ha) until December 2018 and the number of banana lots of the three banana farms in Bocas del Toro, Panama.

Site	Geographic Coordinates		Cultivated Area (ha)	Banana Lots (n)
1	9°25′43.0″ N	82°32′57.1″ W	160.75	56
2	9°28′09.2″ N	82°31′27.9″ W	212.91	77
3	9°30′05.5″ N	82°35′42.2″ W	184.51	47

2.2. Epidemiological Evaluation of Black Sigatoka in Host Plants

The epidemiological parameters to evaluate the incidence and severity of BS were carried out in three ages of plants: plants in the acorn stage, and plants seven weeks (49 days) and ten weeks (70 days) after the appearance of the acorn, in reports of identical format. The only exception is in the parameter called symptom status through the gross sum, which was evaluated only in plants in the acorn stage.

Sampling consisted of weekly monitoring of 10 representative production areas per site. Within each area, 15 plants obtained randomly at the three physiological ages were sampled: five young plants in the acorn stage (1 to 4 days), five plants at seven weeks of age (49 days), and five plants at ten weeks of age (70 days) during the period 2015–2018. The evaluation method based on the severity scale developed by Stover [27], and later modified by Gauhl [28,29], was used. It consisted of visually estimating the affected leaf area, showing necrotic lesions of the stalks go from 4 to 6 on the Fouré scale [30]. The variable related to production was the total number of bunches harvested weekly (52 weeks per year) during the 2015–2018 period for the three sites evaluated (Site 1, Site 2, and Site 3). The epidemiological parameters evaluated were:

a. The infection index (INC): Expresses the magnitude of the damage caused by the disease on a quantitative scale, whose ideal limit is 0.10. It was calculated using Equation (1). This index is a weighted average considering the severity of the BS attack in all the evaluated leaves. The INC gives us a single value indicating the magnitude of foliage damage, on a scale of 0 to 1. In plants in the acorn stage, a value of <0.10 is considered low, ≥ 0.10–0.15 as intermediate, and >0.15 as high [28,29].

$$\mathrm{INC} = \sum \frac{nb}{(\mathrm{N}-1)\mathrm{T} \times 100} \quad (1)$$

where INC = infection index, n = number of leaves in each grade, b = grade, N = number of grades used in the scale, and T = total number of leaves evaluated.

b. State of the symptom (gross sum) (SIN): comes from the observations of the most developed state of the present symptom and the density of these symptoms in sheet No. 3 plants in acorn state. Through a spreadsheet, a numerical value (weight) was obtained for each state of the symptom to subsequently obtain the gross sum. The gross sum reflects the development of BS at an early stage and allows the development of new BS infections to be determined several weeks earlier than with all other parameters. It also allows us to evaluate if the last applied fungicide treatment had a significant effect on the control of BS, or if it is not controlling the disease. The gross sum can range between 0 and 140. A value below 50 is considered good disease control [28,29].

c. The number of functional leaves (LEAF): represents the average of the total number of leaves per plant; only those leaves that, visually, presented 80% of their surface free of necrosis by BS [28,29].

d. Youngest leaf with symptoms (YLWS): number of the leaf-bearing symptoms. Leaves were numbered from the first (topmost) open leaf downward, according to the methodology proposed by Hernández et al. [31]. Youngest leaf with symptoms visible from the ground. These are usually stage 2 and 3 striae that can be identified from the ground.

e. Youngest leaf spotted (YLS): number of the leaf-bearing spots with dry centers. A higher value of YLS indicates more healthy leaves on the plant. Leaves were

numbered as for YLWS. It is the average of the youngest leaf with 10 or more spots with a grayish and dry center, which is the last stage of the development of BS. In the spot state with a grayish and dry center, the fungus produces ascospores [28,29].

These last two variables, YLWS and YLS, represent two concepts of epidemiological importance, such as the incubation period and the latency period. The fluctuations in these two variables are associated with the behavior of the disease in the field, and, therefore, are used as elements of judgment to make decisions in the selection of the fungicide and the days between applications.

2.3. Data Analysis

2.3.1. Exploratory Analysis

The data matrix X was constituted by the set of vectors of the observations $X_{[ij]}$, $j = 1, \ldots, p$, and where each vector $X_{[ij]}$ presented the jth variable for all the observations and where X the data matrix was formed by "$n = 623$" observations with "$p = 13$" variables. A total of 13 predictor variables were considered in this study, being the following: infection index (INC) in plants in the acorn stage; infection index in plants at 7 weeks of age (INC_7) and 10 weeks of age (INC_10); youngest leaf with symptoms on plants at the acorn stage (YLWS) (n°); youngest leaf with symptoms on plants at 7 weeks of age (YLWS_7), and on plants at 10 weeks of age (YLWS_10); youngest leaf spotted in plants in the acorn stage (YLS) (n°), at 7 weeks of age (YLS_7) and 10 weeks of age (YLS_10); more developed state of the BS symptoms in plants in the acorn stage (SIN); and the number of functional leaves in plants in the acorn stage (LEAF) (n°), at 7 weeks of age (LEAF_7) and 10 weeks of age (LEAF_10).

Before data analysis, we checked the data integrity. A principal component analysis (PCA) was performed as an exploratory analysis to check the presence of outliers and identify patterns in the data using the statistical package in R software version 4.0.2 (R Core Team, Vienna, Austria) and the function "PCA" [32]. The objective of this application was to summarize the information of many variables (13 in our study) in a few latent variables, trying to avoid overfitting as new components are added. Pearson's correlation coefficient was also used as a method of parametric statistics, which is not only used to determine the relationship between two quantitative variables, but also to predict a variable.

2.3.2. Random Forest Prediction

The RF algorithm was applied, as a machine learning method based on binary trees, to predict the number of banana bunches. In the scope of our study, RF was used as a regression tool. Predictor variables are evaluated by how much they decreased in node impurity when selected for splits. The impurity of the node is defined as the root mean square error (RMSE) of the node in the RF regression [33].

The statistical program R [32] was used together with the 'randomForest' package with the following configuration (number of variables randomly sampled as candidates at each split (mtry) = 9, number of the trees (ntree) = 500, node size = 5) [34]. Three analysis measures in the package were used: (1) mean decrease accuracy, (2) mean decrease Gini, which are the measures of the performance of the model without each variable, and (3) partial dependence plots.

First, the mean decrease accuracy plot expresses how much accuracy the model loses by excluding each variable. The more precision suffers, the more important the variable is for successful classification. The variables are presented in decreasing importance. The second measure is the mean decrease in the Gini coefficient, which is a measure of how each variable contributes to the homogeneity of nodes and leaves in the resulting random forest. The higher the mean decline precision value or the mean decline Gini score, the greater the importance of the variable in the model [34–36].

The third analysis measure is the partial dependence plots; this shows how each predictor influences the RF model predictions when all other model predictors are controlled. The value of the *Y*-axis of a partial dependence plot is determined by the average of all possible predictions of the model with the data set when the value of the target predictor is *X* [37].

2.3.3. Model Performance Evaluation

The k-fold cross-validation procedure was performed, with k = 3 becoming 3-fold cross-validation. Cross-validation was used primarily in this case to estimate the skill of the machine learning model on unseen data; that is, using a limited sample to estimate how the model is expected to perform overall when used to make predictions on data that was not used during model training.

Four methods were used to evaluate model performance: (1) root mean square error (RMSE), (2) the explanation of variance (%), (3) an observed versus predicted plot to visualize model performance, and the coefficient of determination (R^2). A simple linear regression line was drawn on the plot to compare the accuracy of the model's predictions. These are commonly used measures for farming systems and models of different crops [22,38].

3. Results

3.1. Descriptive Analysis

Table 2 shows the descriptive statistics of the epidemiological parameters and the number of clusters per site evaluated. Site 2 presents the highest average values of the number of banana bunches with 10,855.82 ± 1575.60, and the lowest CV value of 14.5%. On the other hand, site 1 presented the lowest average cluster number value with 7611.77 ± 1648.12 and the highest CV with 21.65%. The asymmetry measures presented served to provide an idea about the shape of the frequency distribution through the coefficient shown in Table 2. The coefficient measures the degree of asymmetry of the distribution concerning the mean. A positive value for this indicator means that the distribution is skewed to the left (positive orientation), as is the case for the banana bunch variables at site 1, YLWS_10, and SIN at all sites. A negative result means that the distribution is skewed to the right, as in the case of the banana bunch variables at sites 2 and 3, and YLS_10 in all the sites evaluated.

Table 2. Summary of descriptive statistics for the three sites evaluated: means values, standard deviations (S.D), standard error (S.E), coefficient of variation, (CV) minimum, maximum, skewness (S), kurtosis (K) and percentiles 25, 50, 75 for the production variable and epidemiological variables.

Site	Variable	Mean	S. D	S. E	CV	Min	Max	S	K	P (25)	P (50)	P (75)
1 (n = 208)	Bunch (nº)	7611.77	1648.12	114.28	21.65	4152.0	12,921.0	0.59	0.25	6404.00	7360.00	8620.00
	INC	0.06	0.03	0.00	41.00	0.02	0.17	1.35	3.09	0.04	0.06	0.08
	YLS (nº)	14.08	1.42	0.10	10.12	9.35	16.60	−1.08	1.09	13.25	14.40	15.16
	YLWS (nº)	11.06	1.41	0.10	12.71	7.45	14.28	−0.08	−0.43	10.10	11.00	11.96
	SIN	52.36	9.91	0.69	18.93	40.00	99.25	1.45	2.62	44.80	49.60	57.60
	LEAF (nº)	13.67	0.65	0.05	4.77	11.50	15.40	−0.47	0.60	13.25	13.70	14.16
	LEAF_7(nº)	10.17	0.89	0.06	8.79	7.25	11.80	−0.62	0.15	9.60	10.28	10.80
	LEAF_10(nº)	9.30	1.20	0.08	12.86	5.15	11.12	−0.76	0.06	8.48	9.48	10.28
	INC_7	0.13	0.05	0.00	42.37	0.03	0.31	1.06	2.02	0.09	0.12	0.16
	YLS_7 (nº)	10.08	1.93	0.13	19.12	4.75	13.20	−0.71	−0.07	8.88	10.48	11.48
	YLWS_7 (nº)	6.43	2.04	0.14	31.73	1.20	10.72	−0.03	−0.22	5.15	6.50	7.56
	INC_10	0.17	0.06	0.00	37.41	0.05	0.39	0.72	0.74	0.13	0.17	0.21
	YLS_10 (nº)	8.66	2.24	0.16	25.86	2.55	12.32	−0.65	−0.27	7.25	9.04	10.24
	YLWS_10 (nº)	4.58	2.08	0.14	45.42	1.00	9.32	0.28	−0.65	3.10	4.32	6.04
2 (n = 208)	Bunch	10,855.82	1575.60	109.25	14.51	6297.00	15,843.00	0.30	0.42	9899.00	10,640.00	11,775.00
	INC	0.08	0.03	0.00	35.06	0.03	0.17	0.85	0.62	0.06	0.07	0.09
	YLS (nº)	13.24	1.93	0.13	14.60	7.95	16.04	−0.78	−0.40	11.96	13.84	14.56
	YLWS (nº)	10.13	1.40	0.10	13.84	6.50	13.36	−0.05	−0.36	9.00	10.28	11.00
	SIN	54.49	10.47	0.73	19.22	40.00	103.00	1.33	2.41	47.20	52.00	59.20
	LEAF (nº)	13.19	1.03	0.07	7.81	10.45	15.12	−0.61	−0.30	12.48	13.36	13.96
	LEAF_7 (nº)	9.47	1.30	0.09	13.73	5.85	11.56	−0.67	−0.53	8.48	9.84	10.52
	LEAF_10(nº)	8.42	1.60	0.11	18.99	4.73	10.92	−0.55	−0.97	7.08	9.00	9.80
	INC_7	0.15	0.06	0.00	40.05	0.06	0.35	1.05	0.85	0.11	0.14	0.18
	YLS_7 (nº)	8.89	2.27	0.16	25.57	3.10	12.56	−0.66	−0.53	7.40	9.40	10.68
	YLWS_7(nº)	5.36	1.94	0.13	36.20	1.00	9.56	−0.19	−0.67	3.80	5.56	6.84
	INC_10	0.21	0.08	0.01	38.67	0.09	0.45	0.90	0.20	0.14	0.19	0.26
	YLS_10 (nº)	7.31	2.51	0.17	34.34	1.00	11.68	−0.75	−0.42	5.85	7.92	9.28
	YLWS_10(nº)	3.53	1.88	0.13	53.35	1.00	7.40	0.22	−1.20	1.70	3.50	5.32

Table 2. *Cont.*

Site	Variable	Mean	S. D	S. E	CV	Min	Max	S	K	P (25)	P (50)	P (75)
3 (n = 207)	Bunch (nº)	8984.02	1479.62	102.84	16.47	5543.00	12,517.00	−0.01	−0.05	8137.00	8969.00	9915.00
	INC	0.07	0.03	0.00	41.15	0.03	0.17	1.17	1.59	0.05	0.06	0.08
	YLS (nº)	13.04	1.49	0.10	11.45	8.45	15.85	−0.94	0.63	12.30	13.36	14.04
	YLWS (nº)	10.53	1.22	0.08	11.59	7.45	13.30	−0.09	−0.40	9.60	10.56	11.35
	SIN	56.76	12.60	0.88	22.19	40.00	106.00	1.17	1.37	48.00	52.80	64.80
	LEAF (nº)	12.89	0.72	0.05	5.61	11.05	14.85	−0.19	−0.30	12.43	12.96	13.40
	LEAF_7 (nº)	9.62	0.75	0.05	7.84	7.45	11.45	−0.66	0.77	9.28	9.70	10.12
	LEAF_10 (nº)	8.73	0.93	0.06	10.70	5.95	10.25	−0.90	0.42	8.30	8.88	9.44
	INC_7	0.14	0.05	0.00	40.45	0.05	0.32	1.17	1.65	0.09	0.13	0.16
	YLS_7 (nº)	9.27	1.78	0.12	19.21	3.78	12.65	−0.81	0.13	8.18	9.72	10.52
	YLWS_7 (nº)	6.07	1.66	0.12	27.36	1.55	9.45	−0.26	−0.28	4.88	6.08	7.28
	INC_10	0.19	0.17	0.01	88.06	0.05	2.43	10.90	137.99	0.13	0.17	0.22
	YLS_10 (nº)	7.85	2.00	0.14	25.48	2.85	11.15	−0.84	−0.23	6.30	8.40	9.36
	YLWS_10 (nº)	4.31	1.74	0.12	40.42	1.00	7.75	0.04	−0.82	3.20	4.12	5.80

Note: Bunch: number of harvested bunches; infection index (INC) in plants in the acorn stage. infection index in plants at 7 weeks of age (INC_7) and 10 weeks of age (INC_10); youngest leaf with symptoms on plants at the acorn stage (YLWS) (nº), youngest leaf with symptoms on plants at 7 weeks of age (YLWS_7) and on plants at 10 weeks of age (YLWS_10); youngest leaf spotted in plants in the acorn stage (YLS) (nº) at 7 weeks of age (YLS_7) and 10 weeks of age (YLS_10); state of the symptom (gross sum) in the acorn stage plants (SIN); the number of functional leaves in plants in the acorn stage (LEAF) (nº) at 7 weeks of age (LEAF_7) and 10 weeks of age (LEAF_10).

On the other hand, some variables present a higher degree of concentration (less dispersion) of the values around their mean, such as the case of LE-AF_10 in sites 1 and 3; and other variables, on the contrary, present a lower degree of concentration (greater dispersion) of their values around their central value, such as the variables YLWS_10 and YLS_10 in all the sites. Therefore, kurtosis informs the pointed (higher concentration) or flattened (lower concentration) of these distributions. Regarding the percentile as a measure of non-central position, for site 2, the P75 of the bunch variable is 9915.0 banana bunches. This means that 75% of the weeks show a bunch production equal to or less than 9915.0, followed by P75 = 9865.0 and 8620.0 banana bunches for sites 3 and 1, respectively.

The correlation analysis is shown as a heat map (Figure 1), useful to try to understand the relationships between the multiple variables studied. The lower triangle of the matrix shows the heatmap of correlations between pairs of variables. The bunch variable has an extremely low correlation coefficient concerning the epidemiological variables; high correlations ($p \leq 0.05$), whose Pearson's r was greater than $\mid 0.85 \mid$ are also shown between the variables derived from the YLS, YLWS, LEAF, and INC.

The PCA was carried out to evaluate the changes in the pattern of the epidemiological variables by evaluated sites. PCA was unable to distinguish the sites evaluated in this study (Figure 2a). The first two principal components (PC) explained 81.6% of the variables; however, no trends were detected in the differences (Figure 2b). According to the value of the contribution in the biplot, we obtained four epidemiological parameters, including YLS_10, INC_7, YLWS_10, and LEAF_7 (Figure 2c). The goodness-of-fit of the PCA model was $R^2X = 0.70$. By analyzing the projections perpendicular to PC1 of the points that represent the cases, certain weeks were identified during the evaluated years with greater inertia, i.e., the points that are at a greater distance from zero (Figure 2c), either moving away to the right or the left. Likewise, the analysis of the projections of the points that represent the variables on the PC1 allowed us to identify the variables with the greatest inertia and with the greatest collinearity, e.g., (YLS10 and YLS7); (LEAF, LEAF 7, and LEAF10); (YLWS, YLWS7, and YLWS10).

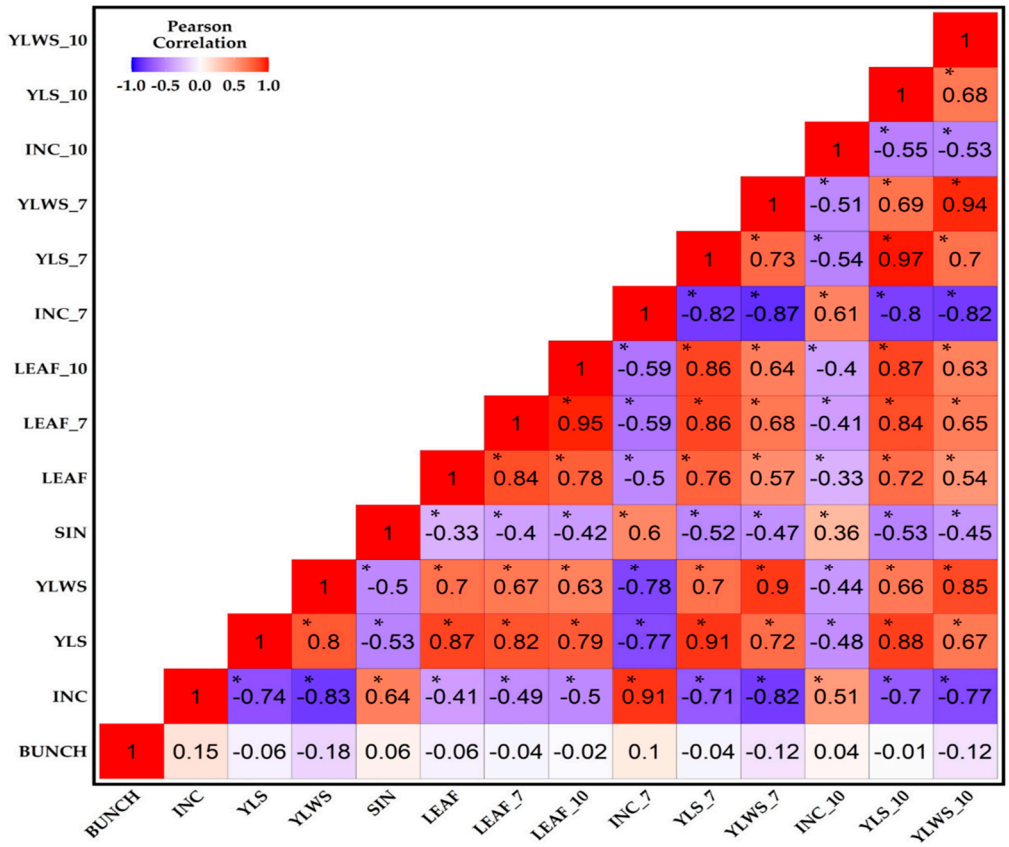


Figure 1. Heatmap of correlation coefficient among epidemiological parameters and numbers of banana bunches. (*) Pearson correlation ($p \leq 0.05$) is shown between variables described in this study. Note: Bunch: number of harvested bunches; infection index (INC) in plants in the acorn stage, infection index in plants at 7 weeks of age (INC_7) and 10 weeks of age (INC_10); youngest leaf with symptoms on plants at the acorn stage (YLWS) (n°); youngest leaf with symptoms on plants at 7 weeks of age (YLWS_7) and on plants at 10 weeks of age (YLWS_10); youngest leaf spotted in plants in the acorn stage (YLS) (n°) at 7 weeks of age (YLS_7) and 10 weeks of age (YLS_10); state of the symptom (gross sum) (SIN); the number of functional leaves in plants in the acorn stage (LEAF) (n°) at 7 weeks of age (LEAF_7) and 10 weeks of age (LEAF_10).

In the interpretation of "correlations" between variables according to the angles of the vectors that represent them, it was found that acute angles indicated positive correlations (all variables derived from YLS, YLWS, and LEAF), obtuse angles corresponded to negative correlations (SIN and INC) and right angles indicated that there was no correlation between the variables. The length of the vectors corresponding to the variables is not of interest when the data has been previously standardized.

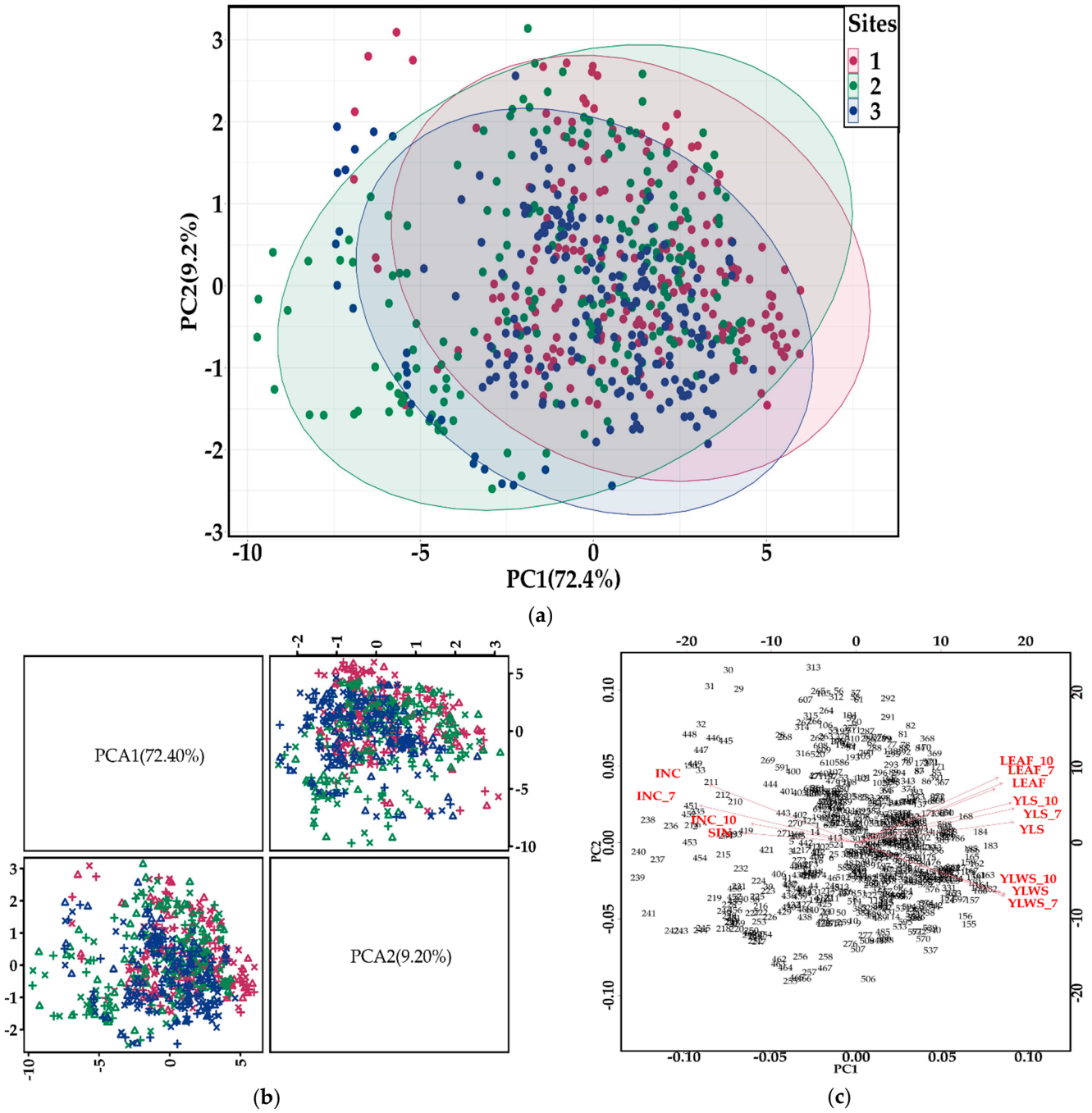


Figure 2. Classifying PC from site based on the epidemiological parameters; (**a**) PCA based on the first two principal components; (**b**) sample scatterplot displays the first two components in each data set in PCA; (**c**) contribution of each feature selected on the first component in PCA using the biplot. Note: Infection index (INC) in plants in the acorn stage, infection index in plants at 7 weeks of age (INC_7) and 10 weeks of age (INC_10); youngest leaf with symptoms on plants at the acorn stage (YLWS) (n°); youngest leaf with symptoms on plants at 7 weeks of age (YLWS_7) and on plants at 10 weeks of age (YLWS_10); youngest leaf spotted in plants in the acorn stage (YLS) (n°) at 7 weeks of age (YLS_7) and 10 weeks of age (YLS_10); state of the symptom (gross sum) (SIN); the number of functional leaves in plants in the acorn stage (LEAF) (n°) at 7 weeks of age (LEAF_7) and 10 weeks of age (LEAF_10).

3.2. Global Banana Bunches Predictions

RF successfully predicted the number of banana bunches on a global scale when compared to test data that had not been included in model training. Figure 1 shows the observed versus predicted plot for the number of banana bunches at all sites tested. The RF model explained 70.6% of the mean yield variation with good agreement between predicted and observed values in the test data (Table 3; Figure 3).

Table 3. Random forest model performance evaluation statistics.

K-Folds	Training Data			Testing Data		
	RMSE (nº Bunches)	Variance Explained (%)	R^2	RMSE (nº Bunches)	Variance Explained (%)	R^2
1	1112.93	69.1	0.69	1132.31	72.0	0.72
2	1112.33	71.0	0.71	1101.23	71.0	0.71
3	1145.67	70.0	0.70	1090.26	69.0	0.69
k-folds CI 95% *	1076–1171	68.0–72.0	0.68–0.72	1054–1162	67.0–74.0	0.67–0.74

* 95% confidence interval (CI).

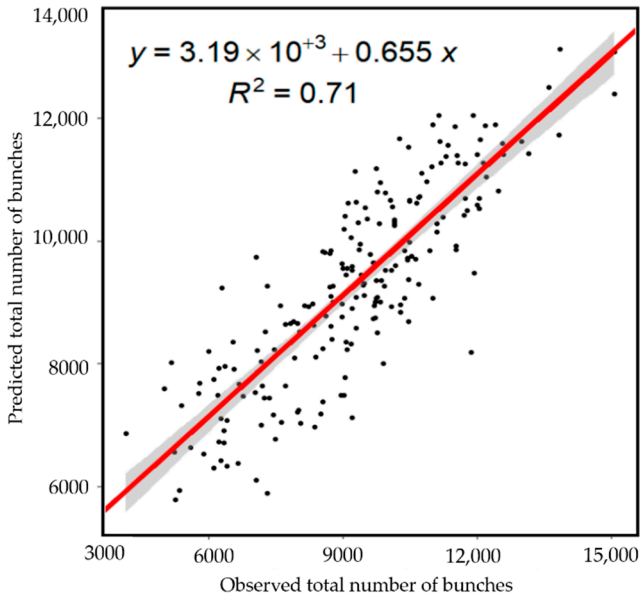


Figure 3. Random Forest model performance for test data sets. The dashed lines indicate a 1:1 relationship and the solid line represents the linear regression between the observations and the predictions made for the test data sets.

RMSE values were within the 95% confidence interval of 1054–1162 banana bunches, which is 11.51% of the observed average yield. The comparison between the number of banana bunches observed and predicted (Figure 3) indicates that the predictions of the RF model are in good agreement with the observations with a slope of 0.65 and the R^2 of 0.71.

The RF method was performed with Mtry = 9, ntree = 500; the OOB (out-of-bag) error was found to be minimal, as shown in Figure 4a. About two-thirds of the sample, called the in-bag sample, was used to train the model, and the remaining third of the sample, called the out-of-bag sample, was used for internal cross-validation on the RF model.

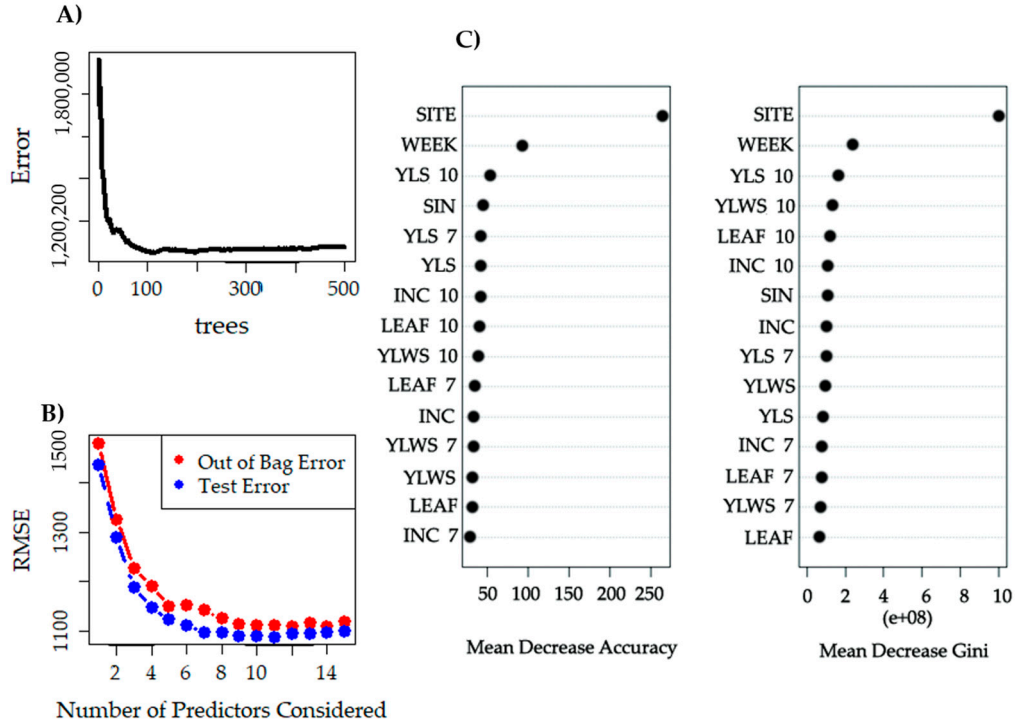


Figure 4. RF estimation for the number of banana bunches: (**A**) OOB error, (**B**) predictor variable importance graph, and (**C**) selection of a few variables using 3-fold cross-validation based on least RMSE.

The prediction of the number of banana bunches was based on the 13 independent variables on a global scale over four years. The main predictor variables in the graph of the importance of the variable obtained from the RF are shown in Figure 4b. The influence of a set of six variables for the prediction of the number of banana bunches is observed through the recursive elimination of characteristics with the cross-validation technique, as shown in Figure 4c.

Additionally, RF can also provide useful information about the importance and dependence of the variable on its predictive capacity. The range of importance of the epidemiological variables and the partial impact of the variable on the response represented by the number of banana bunches can be evaluated for systems analysis purposes. The use of the measure of variable importance, represented by mean decrease accuracy and mean decrease Gini, allowed us to identify the most influential variables that determine banana production in the evaluated sites that we tested (Figure 4b).

Partial dependence plots were useful to assess the relationship between each predictor and the response variable (Figure 5), measure the effect of explanatory variables, such as site (Figure 5a), week (Figure 5b), SIN (Figure 5c), YLS_10 (Figure 5d) and YLWS_10 (Figure 5e), on the variable number of banana bunches. Variable importance measures and partial dependence plots revealed unique responses concerning the evaluated variables. The RF model identified the trend of the YLS parameter in plants with 10 weeks of physiological age, being one of the most influential variables responsible for the increase in the number of bunches (>1000) during the four years; the ideal values of this parameter are between blade position 10 and 12 (Figure 5d). The site was ranked as the first and most important variable for estimating banana yield (Figure 5a). Likewise, the week was classified as the second most important variable for the number of banana bunches, since the yield or the banana bunches fell in weeks 18 to 22, which indicates that the week of the year can be a key factor to achieve a high yield of bananas in this site, based on the evaluation of the BS parameters (Figure 5b). For example, the partial dependence plot shows that the epidemiological parameters such as SIN were one of the main predictors of the number of banana bunches, with a saturation response at SIN values less than 40 (Figure 5c). The

partial dependence graphs for other variables studied presented minor effects with very complex patterns that are difficult to explain through certain physiological or agronomic responses of bananas.

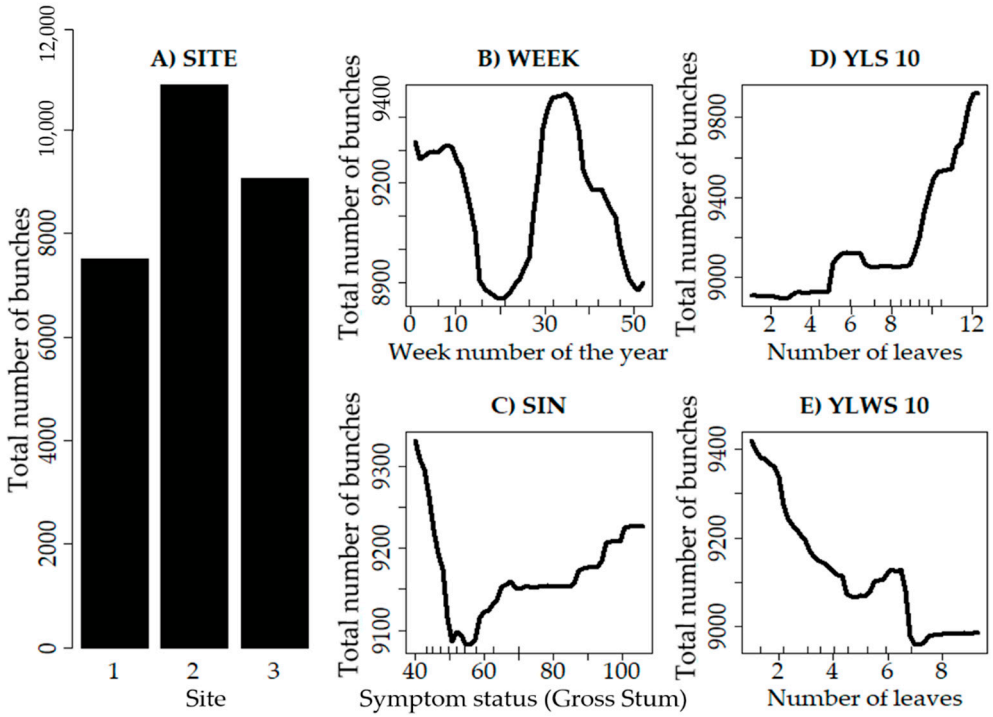


Figure 5. Partial dependence plots for the top-ranked predictor variable from variable importance measures of random forest models: (**A**) site, (**B**) week, (**C**) SIN, (**D**) YLS_10, and (**E**) YLWS_10. The Y-axis of each plot indicates the average of all the possible model predictions for the *X* predictor value. The *X*-axis hash marks indicate deciles.

3.3. Identification of the Main Variables

The behavior of the disease was subject to a dynamic that depended on environmental conditions during the 2015–2018 period, which showed different phases of development, indicating that the disease is cyclical, showing variable fluctuations during the year and between years. In terms of evaluation, sharp exponential growth was seen in the month of August weeks 32–35 according to the SIN, with 2015 being the year with the highest reported values of SIN (99.25–106.0) for all sites (Appendix A Figure A1(1–3)). Then, a phase of growth slowdown manifested itself in August 2016.

For the evaluated sites, a generalized trend of the maximum number of bunches harvested globally was observed, located in weeks 32 to 36 with values higher than 11,700 bunches for the period 2017–2018 in site 1; for site 2, it was evidenced that higher values of the number of clusters were obtained in weeks 32 to 35, being higher than 15,000; and finally in site 3, the maximum number of clusters harvested was during weeks 30–33, with values greater than 12,000 bunches. This establishes a clear pattern that the number of the week, together with the evaluation of epidemiological parameters, have an important weight for the management of BS and production. The YLWS_10 presented a curve with maximums in three seasons during the years 2017 and 2018, with week 4/2018 corresponding to 22–28 January, being the highest value (9.32) for site 1, followed by week 26/2018, corresponding to the 25 June to 1 July, with 9.16 for site 1 and 7.75 for site 3, which indicates that the higher the value of YLWS_10, the slower the development of the disease. On the contrary, low values of YLWS_10, such as those recorded during weeks 31–36/2015 correspond to the end of July and the month of August, which indicate that the development of the disease is faster during that time.

Regarding the YLS_10 in general, it was evidenced that the maximum values of this parameter were in weeks 5/2018 and 7/2018, corresponding to the first fortnight of February with values of 11.68 and 12.30, respectively (Appendix A Figure A1(1–3)); likewise, all the sites presented maximum values of YLS_10 during the weeks 30/2018 (10.64) and 33/2018 (11.8), corresponding to the days of 23 to 27 July and August 13 to 19. The lowest values of this parameter were in weeks 32, 33, and 35 of 2015, corresponding to the month of August, whose values fluctuated between 1.0–2.95.

In the case of the variable LEAF_10, all the sites showed a tendency to a considerable reduction in the number of functional leaves during weeks 34–38/2015, corresponding to the month of August and mid-September, with values that ranged between 5 and 6 functional leaves. Site 2 presented the highest concentration of functional leaves below 8 leaves in 2018, with values of 5 functional leaves (Appendix A Figure A1(1–3)).

The SIN variable presented its maximum points mostly during week 32/2015, corresponding to 3–9 August, for all sites evaluated with values of 103.0–106.0. On the other hand, the lowest values were located mostly in week 17/2017, corresponding to 24–30 April, whose values were 40.0. However, the trend of the curve is more important than the absolute value of the SIN. In all cases, the SIN continued to rise three weeks in a row, entering critical levels (>50) (Appendix A Figure A1(1–3)); therefore, a careful review of the fungicide application program is necessary for the weeks before and after these levels, since it is an indicator that there is a high potential for a strong infection in the plantation.

The time when the symptoms of the disease manifested with greater severity was in the months of July, August, and part of September in most years; after this time, it can be said that a self-destruction phase of the disease was defined. Fungus, in the months of December–January, and May–July; since a recovery period of the plant could be observed, corroborated by the increase in the values of the YLWS, YLS, and LEAF (Appendix A Figure A1(1–3)).

Once the dynamics of the disease are known, an integrated management program for BS can be established, according to the results obtained, the phytosanitary control practices in the evaluated area and areas with similar climatic conditions, in the months of growth or exponential phase of the disease. In this area, it is appropriate to carry out preventive control in April, June–July, and November.

4. Discussion

In the process of filling the fruits of banana cultivation, the progressive development of the leaf area and its physiological activity are determining factors in the productive performance, for which it is essential that, during its growth, the plant has enough functional leaves, which guarantee that photosynthesis is carried out optimally [39]. In this sense, to obtain a good production of banana bunches, a minimum of eight functional leaves must be maintained at the time of flowering [40]. The foregoing indicates the importance of agronomic management to reach flowering, at least, with eight healthy leaves free of BS [41,42].

The mean lowest number of functional leaves was found in site 2 with 4.73, indicating that the management conditions and even the environmental conditions in the plantation of this site could provide the crop with a negative effect on the number of healthy functional leaves. In this regard, a study showed that greater health in the plantation is determined by a greater number of functional leaves when the crop is grown in an environment of 20% shade [42]. These results coincide with the investigations of Belalcázar et al. [43] which indicate that it is necessary to establish a simultaneous agroforestry system with banana cultivation at distances that do not interfere directly with the plants, such as, for example, 4 m between rows of plants and 2 m between plants in the row, since it helps to reduce the severity of BS disease.

The research reported by Barrera et al. [41] established that banana plants that kept between eight and twelve leaves during flowering had larger bunches, so to guarantee a good filling the plant at harvest must end with at least six functional leaves. Additionally,

the results of studies, such as those by Cayón [44] and Siles et al. [45], describe that the highest rates of photosynthesis occur in the youngest leaves of the plant (two, three, four, and five), drastically reducing in leaves six, seven, eight and nine; these are older leaves, which shows the possibility that certain microclimatic conditions of shade in early stages of plant development may favor the crop, for its general vigor and better response to the appearance of BS.

Bananas shed between 0.8 and 1.2 leaves each week. From this, it is inferred that an important part of the symptoms that are observed weekly in the fourth leaf come from inoculations that occurred during the different stages of opening of the leaves or when they are just opened [46]. Likewise, in Costa Rica with a climate like that of our study, the symptoms took 9.6 days more to develop after flowering of the False Horn banana cv. 'Currare' than before flowering [47]. These results suggest that the host's response to BS depends on the host's metabolism. The difference in host response observed between the vegetative and generative stages of plantain could be due to changes in the hormonal system of the central meristem that starts to flower and simultaneously stops the formation of new leaves [48].

In leaves affected by BS, photosynthesis and respiration processes are affected, causing irreparable morphological and physiological damage, which prevents the proper functioning of the crop, directly damaging its development and production [7,49], which generates, therefore, low productivity, premature ripening of fruits and a decrease in quality [50].

Banana plants were more sensitive to BS, presenting the YLWS between leaves 1 and 4 (Table 2), which are below the number of leaves necessary to develop a good bunch, affecting yield and the quality of the fruit. The YLWS provides information on the quality of control in young leaves and the BS progress in the plantation [49].

YLWS is a widely used variable, captured weekly by inspectors from the plant health departments of aerial spraying companies. This parameter presented an important influence in the RF model and in a certain way it has a good relationship with the production of bunches. That is, it is possible to predict the future level of the number of banana bunches knowing the status of this epidemiological parameter, requiring in general terms a greater number of variables and especially the evaluation of their interaction.

The study by Cedeño et al. [51] established that, with defoliation and surgery at weekly intervals, the severity of the disease was reduced. Other practices in addition to sanitary defoliation could also help in its management, such as piling or cordoning off diseased tissue on the ground and the application of 10% urea as a sporulation inhibitor [52].

The rank of the youngest spotted leaf (counted from the last emerging leaf), called YLS, is an especially important index that provides information on the growth rate of necrotic lesions responsible for the destruction of large portions of leaf tissue. The lower the YLS, the faster the development of the disease. The YLS_10 was on average between leaf number seven and eight; ideally, it would be between leaf number ten and eleven, indicating that the banana plant can reach the end of its production cycle, which would allow a good filling of the fruits and, therefore, high production of bunches.

When a younger diseased leaf is detected in the field among the first leaves of the plant, obtaining good yields and bunch production is aggravated, since within the agronomic management of BS, if it is decided to carry out defoliation operations and pruning to reduce the more rapid ascent of the disease towards the upper leaves, all old and infected leaves should be carefully cut before flowering, which would result in several leaves below six, the minimum quantity necessary to avoid a decrease bunch production [53].

In the reproductive cycle of the banana plant (*Musa* AAB cv Harton) from the exit of the bunch to its harvest, 70 to 98 days (12 $\pm$ 2 weeks) pass under the conditions of the banana zone of the South of Lake Maracaibo, Venezuela [49]. The reports of this last study indicate that the minimum number of leaves present at the beginning of the cycle should be 12 to 13 in bananas. It is understood that the three upper leaves (the youngest) supply the needs of the plant, and older leaves help bunch growth.

In this regard, in the climate of the South Pacific of Mexico, Black Sigatoka showed that the greatest damage (12 to 25% severity) occurs during the months of June to December, which coincides with the season of greatest rainfall [54]. In this period, symptoms appear in a state of spot between leaf nos. 4 to 6 and 25 to 58% of diseased leaves. The lower severity of the disease (January to May) in the South Pacific is related to the period of lower rainfall [55], where spots appear between leaves No. 7 to 9 and 7 to 25% diseased leaves [56].

Unfortunately, this disease reduces leaf longevity and threatens the plant's ability to achieve full fruit fill. In practical terms, it is recommended to use the variables YLWS, YLS, and LEAF to evaluate the seasonal behavior of the disease at the farm level. Although the analysis methods may be complex for technicians in the region, at a later stage, it is recommended to implement, through an algorithm such as the one developed in this study, all the analysis procedures required for modeling the disease, in such a way that a technician only has to enter the data from the monitoring of BS to make adjustments to the models and, through a platform, predict the future evolution of bunch production.

The results generated in this study emphasize the obtaining of scientific information from commercial plantations, which can be complemented by the addition of data and information on cultural practices and chemical control carried out, applications of biological inputs, such as promoters of growth and resistance to disease, or biological regulators of pathogen populations, among other practices, which could increase the prediction accuracy of the RF model by reducing the mean square error of prediction, which would finally translate in timely applications of fungicides (protectants) and not calendar applications as is currently done, thus generating a reduction in the number of cycles of systemic fungicides by being used only when there is an increase in the evolution of the disease that justifies it.

The models developed, such as the RF model, can become a planning tool for production and disease management in terms of cycles, product rotation, and times to use the different types of products. On the other hand, as more farms adopt the methodology, the spraying frequencies will probably not be the same for all sites or farms.

The decision to use RF for this type of problem was associated with the main advantage of RF regression when the explanatory variables are highly correlated (e.g., the number of functional sheets, YLS, and YLWS at different stages, and physiological characteristics of the plant) [57]. Many predictors of crop production, such as these types of variables that are repeated in the evaluation, are often highly correlated with each other and can have multicollinearity. RF uses the best single variable when splitting the responses at each node of the decision trees and averaging the predictions from the trees in the forest to make a multidimensional step function [36]. This means that even if several variables are correlated and drive the response similarly, only one of them can affect the RF regression model at a time [38].

The results demonstrate that RF regression is highly effective for farm-scale crop yield predictions. Although RF has been widely used as a classification algorithm for various applications recently [11,21], nowadays, few studies have analyzed its regression capabilities for banana production in Latin American territories. Our results are shown here to establish that the RF regression has significant merits that are highly suitable for predicting the production of crops, such as bananas with SB epidemiological parameters.

Finally, a wide range of factors that affect Sigatoka has been reported in the literature; for example, shade, soil type, drainage, use of fertilizers, plant density, and irrigation [8,14], as well as the mineral nutrition of plants, considered as an exogenous factor, which can be modified, and constitutes an additional fundamental point to combat diseases [11,13]. Any reduction in the number of bunches harvested or in the weight of the bunch in Musaceae could be the effect of the low initial number of leaves and/or their deterioration due to the action of pathogens or other biotic agents. In the case of BS attacks, the economic critical point must be established to proceed with the control of the disease, and this value is related to the effective leaf area over time.

5. Conclusions

This study evaluated the effectiveness of banana bunch prediction of the RF algorithm, which has certain advantages for the regression of crop systems such as bananas but is not yet widely used in this field. We show that RF provides a production based on epidemiological parameters of black sigatoka. The result of this study shows great potential for the use of the RF algorithm with the accuracy of the model's predictions of 0.71, so RF is an alternative statistical modeling method for production predictions in bananas.

The information on the prediction of the production of banana bunches can help the producer to estimate the profitability of the harvest and encourage the generation and monitoring of studies to verify which epidemiological parameters of black sigatoka most influence the final production. In addition, one of the biggest problems in the management of this disease is the time or incubation period and the latency period of black sigatoka, represented by the youngest leaf with symptoms and youngest leaf spotted parameters, being important elements of judgment in decisions related to the selection of the fungicide and the days between applications. Therefore, any tool that reduces this period will be of great help.

In summary, the results support that RF regression may be a suitable algorithm to predict the number of banana bunches at such sites with careful selection of a training dataset that includes a diverse range of predictors, with the site, week, youngest leaf spotted, and youngest leaf with symptoms in plants with 10 weeks of physiological age as the best predictor groups.

Author Contributions: Conceptualization, B.O.O., M.A.R.C. and A.V.; methodology, M.A.R.C. and A.V.; software, B.O.O., M.A.R.C., M.A.-A. and A.V.; validation, B.O.O., M.A.R.C. and A.V.; formal analysis, M.A.R.C. and A.V.; investigation, B.O.O., M.A.-A., E.M. and E.M.-G.; resources, B.O.O., M.A.-A., E.M. and E.M.-G.; data curation, B.O.O. and E.M.-G.; writing—original draft preparation, B.O.O.; writing—review and editing, M.A.-A., E.M. and E.M.-G.; visualization, M.A.-A., M.A.R.C. and A.V.; supervision, M.A.-A., E.M.-G. and E.M.; project administration, E.M.-G.; funding acquisition, B.O.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are grateful for the international mobility scholarship from the Campus of International Excellence for the Environment, Biodiversity and Global Change (CEI-Cambio) based in Seville, Spain; for promoting outreach activities in indigenous territories of Panama. Additionally, to the Board of Directors of the Cooperativa Bananera del Atlántico (COOBANA RL), the Management, collaborators, and technical personnel, especially Ing. Ernesto Ortiz (Phytosanitary Technical Advisor), Luis Escobar (Sigatoka Technical Supervisor), Juan Corella, and Diomedes Rodríguez, for his valuable support in obtaining epidemiological information, and that will serve as a guide for decision-making in the sustainable production of bananas for export by COOBANA, RL.

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

Appendix A

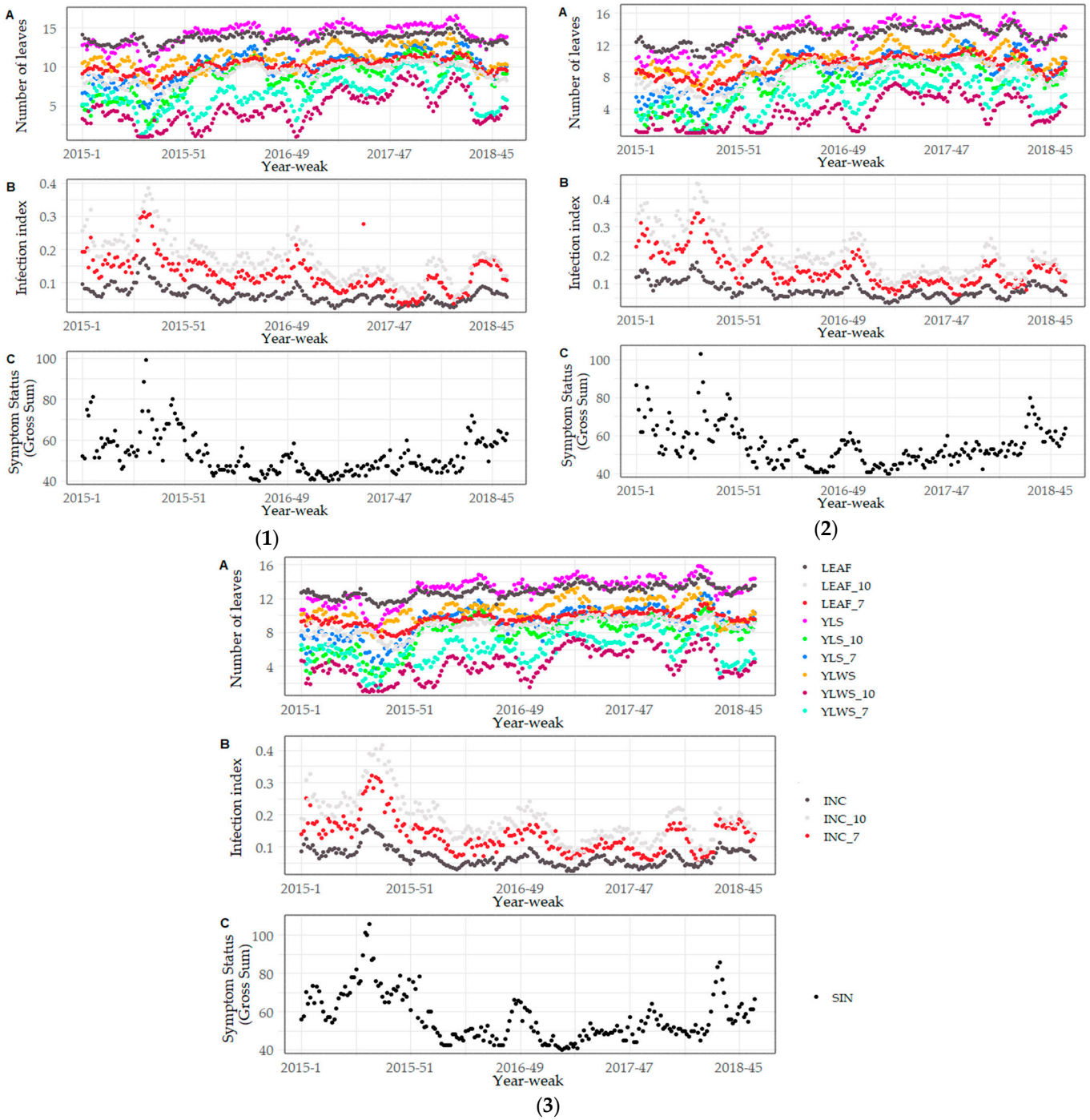


Figure A1. Weekly distribution of epidemiological parameters (2015–2018) at site 1 (**1**), site 2 (**2**), and site 3 (**3**). (**A**) Youngest leaf with symptoms on plants at the acorn stage (YLWS) (nº), youngest leaf with symptoms on plants at 7 weeks of age (YLWS_7) and on plants at 10 weeks of age (YLWS_10); youngest leaf spotted in plants in the acorn stage (YLS) (nº) at 7 weeks of age (YLS_7) and 10 weeks of age (YLS_10); the number of functional leaves in plants in the acorn stage (LEAF) (nº) at 7 weeks of age (LEAF_7) and 10 weeks of age (LEAF_10). (**B**) Infection index (INC) in plants in the acorn stage, infection index in plants at 7 weeks of age (INC_7) and 10 weeks of age (INC_10). (**C**) State of the symptom (gross sum) in the acorn stage plants (SIN).

References

1. Food and Agriculture Organization of the United Nation. Banana Market Review: Preliminary Results 2020. Available online: https://www.fao.org/3/cb5150en/cb5150en.pdf (accessed on 16 September 2022).
2. Evans, E.A.; Ballen, F.H.; Siddiq, M. Banana production, global trade, consumption trends, postharvest handling and processing. In *Handbook of Banana Production, Postharvest Science, Processing Technology and Nutrition*; Siddiq, M., Ahmed, J., Lobo, M.G., Eds.; Wiley Online Library: New York, NY, USA, 2020; pp. 1–18. [CrossRef]
3. Pitti, J.; Olivares, B.O.; Montenegro, E.; Miller, L.; Ñango, Y. The role of agriculture in the Changuinola district: A case of applied economics in Panama. *Trop. Subtrop. Agroecosyst.* **2021**, *25*, 017. [CrossRef]
4. Martínez-Solórzano, G.E.; Rey-Brina, J.C. Bananos (Musa AAA): Importance, production, and trade in COVID-19 times. *Agron. Mesoam.* **2021**, *32*, 1034–1046. [CrossRef]
5. Montenegro, E.J.; Pitti-Rodríguez, J.E.; Olivares-Campos, B.O. Identification of the main subsistence crops of Teribe: A case study based on multivariate techniques. *Idesia* **2021**, *39*, 83–94. [CrossRef]
6. Rhodes, P.L. A new Banana disease in Fiji. *Commonw. Phytopathol. News* **1964**, *10*, 38–41.
7. Marin, D.H.; Romero, R.A.; Guzman, M.; Sutton, T.B. Black Sigatoka: An increasing threat to banana cultivation. *Plant Dis.* **2003**, *87*, 208–222. [CrossRef] [PubMed]
8. Fullerton, R.A. Sigatoka Leaf Diseases. In *Compendium of Tropical Fruit Diseases*; Ploetz, R.C., Zentmyer, G.A., Nishijinia, W.T., Rohrbach, K.G., Ohr, H.D., Eds.; American Phytopathological Society: St. Paul, MN, USA, 1994; pp. 12–14.
9. Jacome, L.H.; Schuh, W.; Stevenson, R.E. Effect of temperature and relative humidity on germination and germ tube development of *Mycosphaerella fijiensis* var. *difformis*. *Phytopathology* **1991**, *81*, 1480–1485. [CrossRef]
10. Fullerton, R.A.; Casonato, S.G. The infection of the fruit of 'Cavendish' banana by *Pseudocercospora fijiensis*. cause of black leaf streak (black Sigatoka). *Eur. J. Plant Pathol.* **2019**, *155*, 779–787. [CrossRef]
11. Olivares, B.O.; Vega, A.; Calderón, M.A.R.; Rey, J.C.; Lobo, D.; Gómez, J.A.; Landa, B.B. Identification of Soil Properties Associated with the Incidence of Banana Wilt Using Supervised Methods. *Plants* **2022**, *11*, 2070. [CrossRef] [PubMed]
12. Dita, M.; Barquero, M.; Heck, D.; Mizubuti, E.S.; Staver, C.P. Fusarium wilt of banana: Current knowledge on epidemiology and research needs toward sustainable disease management. *Front. Plant Sci.* **2018**, *9*, 1468. [CrossRef]
13. Olivares, B.O.; Paredes, F.; Rey, J.C.; Lobo, D.; Galvis-Causil, S. The relationship between the normalized difference vegetation index. rainfall. and potential evapotranspiration in a banana plantation of Venezuela. *ST-JSSA* **2021**, *18*, 58–64. [CrossRef]
14. Churchill, A.C. *Mycosphaerella fijiensis*. the black leaf streak pathogen of banana: Progress towards understanding pathogen biology and detection. disease development. and the challenges of control. *Mol. Plant Pathol.* **2011**, *12*, 307–328. [CrossRef]
15. Wang, Y.; Chee, M.C.; Edher, Z.; Hoang, M.D.; Fujimori, S.; Kathirgamanathan, S.; Bettencourt, J. Forecasting black sigatoka infection risks with latent neural odes. *arXiv* **2020**, arXiv:2012.00752. [CrossRef]
16. Ochoa, A.; Abaunza, F.; Rey, V. Forecasting black sigatoka in banana crops with stochastic models. In Proceedings of the VI International Banana Congress CORBANA, Miami, FL, USA, 19–20 April 2016.
17. Ganry, J.; de Bellaire, L.D.L.; Mourichon, X. A biological forecasting system to control Sigatoka disease of bananas and plantains. *Fruits* **2008**, *63*, 381–387. [CrossRef]
18. de Hernández, R.M.A.G.; Rodríguez, V.; Caraballo, E.A.H. Predicción del rendimiento de un cultivo de plátano mediante redes neuronales artificiales de regresión generalizada. *Publ. Cienc. Tecnol.* **2012**, *6*, 31–40.
19. Soares, J.D.R.; Pasqual, M.; Lacerda, W.S.; Silva, S.O.; Donato, S.L.R. Utilization of artificial neural networks in the prediction of the bunches' weight in banana plants. *Sci. Hortic.* **2013**, *155*, 24–29. [CrossRef]
20. Ye, H.C.; Huang, W.J.; Huang, S.Y.; Cui, B.; Dong, Y.Y.; Guo, A.T.; Ren, Y.; Jin, Y. Identification of banana fusarium wilt using supervised classification algorithms with UAV-based multi-spectral imagery. *Int. J. Agric. Biol.* **2020**, *13*, 136–142. [CrossRef]
21. Selvaraj, M.G.; Vergara, A.; Montenegro, F.; Ruiz, H.A.; Safari, N.; Raymaekers, D.; Ocimati, W.; Ntamwira, J.; Tits, L.; Omondi, A.B.; et al. Detection of banana plants and their major diseases through aerial images and machine learning methods: A case study in DR Congo and Republic of Benin. *ISPRS* **2020**, *169*, 110–124. [CrossRef]
22. Olivares, B.O.; Araya-Alman, M.; Acevedo-Opazo, C.; Rey, J.C.; Cañete-Salinas, P.; Kurina, F.G.; Balzarini, M.; Lobo, D.; Navas-Cortés, J.A.; Landa, B.B.; et al. Relationship between soil properties and banana productivity in the two main cultivation areas in Venezuela. *J. Soil Sci. Plant Nutr.* **2020**, *20*, 2512–2524. [CrossRef]
23. Sangeetha, T.; Lavanya, G.; Jeyabharathi, D.; Rajesh Kumar, T.; Mythili, K. Detection of pest and disease in banana leaf using convolution Random Forest. *Test Eng. Manag.* **2020**, *83*, 3727–3735.
24. Hou, J.C.; Hu, Y.H.; Hou, L.X.; Guo, K.Q.; Satake, T. Classification of ripening stages of bananas based on support vector machine. *Int. J. Agric. Biol. Eng.* **2015**, *8*, 99–103. [CrossRef]
25. Sabilla, I.; Wahyuni, C.S.; Fatichah, C.; Herumurti, D. Determining banana types and ripeness from image using machine learning methods. In Proceedings of the 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT), Ouargla, Algeria, 13–15 March 2019; IEEE: New York, NY, USA, 2019; pp. 407–412.
26. Montenegro, E.; Pitti, J. Analysis of climate risks in the indigenous food system The Teribe, Panama. *Acta Nova* **2020**, *9*, 713–736.
27. Stover, R.H. A proposed international scale for estimating intensity of Banana leaf spot (*Mycosphaerella musicola* Leach). *Trop. Agric. Trinidad Tobago* **1971**, *48*, 185–196.
28. Gauhl, F. *Epidemiología y Ecología de la Sigatoka Negra (Mycosphaerella fijiensis, Morelet) en Plátano (Musa sp.), en Costa Rica*; Trad; de Alemán, J.E., Ed.; UPEB: Panama City, Panama, 1990; p. 126.

29. Gauhl, F.; Pasberg-Gauhl, C.; Jones, D.R. Disease cycle and epidemiology. In *Diseases of Banana. Abacá and Enset*; Jones, D.R., Ed.; CAB International: Wallingford, UK, 2000; pp. 56–62.
30. Fouré, E. *Black Leaf Streak Disease of Bananas and Plantains (Mycosphaerella fijiensis Morelet). Study of the symptoms and Stages of the Disease in Gabon*; IRFA-CIRAD: Paris, France, 1985; p. 20.
31. Hernández, L.; Hidalgo, W.; Linares, B.; Hernández, J.; Romero, N.; Fernández, S. Surveillance and forecast preliminary study for black sigatoka (*Mycosphaerella fijiensis* Morelet) disease in *Musa* AAB cv Hartón plantain crop in "Macagua-Jurimiquire" Yaracuy state. *Rev. Fac. Agron.* **2005**, *22*, 325–329.
32. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
33. Breiman, L.; Last, M.; Rice, J. Random forests: Finding quasars. In *Statistical Challenges in Astronomy*; Feigelson, E., Jogesh, G., Eds.; Springer: New York, NY, USA, 2003; pp. 243–254. [CrossRef]
34. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
35. Strobl, C.; Boulesteix, A.L.; Zeileis, A.; Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources, and a solution. *BMC Bioinform.* **2007**, *8*, 1–25. [CrossRef] [PubMed]
36. Han, H.; Guo, X.; Yu, H. Variable selection using mean decrease accuracy and mean decrease gini based on random forest. In Proceedings of the 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 26–28 August 2016; pp. 219–224. [CrossRef]
37. Liaw, A.; Wiener, M. Classification, and regression by random Forest. *R News* **2002**, *2*, 18–22.
38. 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] [PubMed]
39. Arcila, M.; Giraldo, G.; Duarte, J. *Influencia de Las Condiciones Ambientales Sobre Las Propiedades Físicas y Químicas Durante La Maduración del Fruto de Plátano Dominico-Hartón (Musa AAB Simonds) en la Zona Cafetera Central*; Cayón, G., Ed.; Poscosecha y Agroindustria del Plátano en el eje Cafetero de Colombia; Universidad del Quindío: Quindío, Colombia, 2000; pp. 101–124.
40. Torres, N.; Hernández, J. Efecto del número de hojas en el desarrollo del racimo de plátano Hartón *Musa* AAB. *Agroalim. Des. Sost.* **2004**, *5*, 17–22.
41. Barrera, J.; Cayón, G.; Robles, J. Influence of leaf and fruit epicarp exposition on development and quality of 'Hartón' plantain (*Musa* AAB Simmonds) bunch. *Agron. Colomb.* **2009**, *27*, 73–79.
42. Barrera, V.J.; Barraza, A.F.; Campo, A. Shadow effect on black sigatoka (*Mycosphaerella fijiensis* Morelet) in plantain cultivation cv harton (*Musa* AAB Simmonds). *Rev. UDCA Actual. Divulg. Cient.* **2016**, *19*, 317–323. [CrossRef]
43. Belalcázar, S.; Merchán, V.; Mayorga, M. *Control de Enfermedades. El Cultivo Del Plátano en el Trópico*; Belalcázar, S., Ed.; Feriva: Bogota, Colombia, 1991; pp. 241–297.
44. Cayón, G. Evolución de la fotosíntesis, transpiración y clorofila durante el desarrollo de la hoja de plátano (*Musa* AAB Simmonds). *Infomusa* **2001**, *10*, 12–15.
45. Siles, P.; Bustamante, O.; Valdivia, E.; Burkhardt, J.; Staver, C. Photosynthetic performance of banana (Gross Michel, AAA) under a natural shade gradient. *Acta Hortic.* **2013**, *986*, 71–77. [CrossRef]
46. Porras, A.; Hernández, A.; Pérez, L. Epidemiología de la Sigatoka Negra (*Mycosphaerella fijiensis* Morelet) en Cuba. II. Pronóstico Bio-Climático de los Tratamientos contra la Enfermedad en Plátanos (*Musa* spp. AAB). *Rev. Mex. Fitopatol.* **2000**, *18*, 27–35.
47. Gauhl, F.; Pasberg-Gauhl, C. Epidemiology of black sigatoka disease on plantain in Nigeria. *Phytopathology* **1994**, *84*, 1080.
48. Mobambo, K.N.; Gauhl, F.; Pasberg-Gauhl, C.; Zuofa, K. Season and plant age effect evaluation of plantain for response to black sigatoka disease. *Crop Prot.* **1996**, *15*, 609–614. [CrossRef]
49. Nava, C.; Vera, J. Relation of leaves numbers a florewing time and wasted at reproductive cycle with bunch weight in plantain bunch plants under black Sigatoka attack. *Rev. Fac. Agron.* **2004**, *21*, 335–342.
50. Rodríguez, P.; Cayón, G. Efecto de *Mycosphaerella fijiensis* sobre la fisiología de la hoja de banano. *Agron. Colomb.* **2008**, *26*, 256–265.
51. Cedeño-Zambrano, J.R.; Díaz-Barrios, E.J.; de Jesús Conde-López, E.; Cervantes-Álava, A.R.; Avellán-Vásquez, L.E.; Zambrano-Mendoza, M.E.; Sánchez-Urdaneta, A.B. Evaluación de la severidad de Sigatoka negra (*Mycosphaerella fijiensis* Morelet) en platano "Barraganete" bajo fertilización con magnesio. *Rev. Tec.* **2021**, *44*, 4–12. [CrossRef]
52. Villalta, R.; Guzmán, M. Capacidad de esporulación de *Mycosphaerella fijiensis* en tejido foliar de banano depositado en el suelo y efecto antiesporulante de la urea. *Corbana* **2005**, *31*, 41–43.
53. Etebu, E.; Young, W. Control of black sigatoka disease: Challenges and prospects. *Afr. J. Agric. Res.* **2011**, *6*, 508–514. [CrossRef]
54. Vázquez-Euán, R.; Chi-Manzanero, B.; Hernández-Velázquez, I.; Tzec-Simá, M.; Islas-Flores, I.; Martínez-Bolaños, L.; Garrido-Ramírez, E.R.; Canto-Canché, B. Identification of new hosts of *Pseudocercospora fijiensis* suggests innovative pest management programs for black sigatoka disease in banana plantations. *Agronomy* **2019**, *9*, 666. [CrossRef]
55. Gómez-Correa, J.C.; Torres-Aponte, W.S.; Cayón-Salinas, D.G.; Hoyos-Carvajal, L.M.; Castañeda-Sánchez, D.A. Modelación espacial de la Sigatoka negra (*Mycosphaerella fijiensis* M. Morelet) en banano cv. Gran Enano. *Rev. Ceres* **2017**, *64*, 47–54. [CrossRef]
56. Bebber, D.P. Climate change effects on Black Sigatoka disease of banana. *Philos. Trans. R. Soc. B* **2019**, *374*, 20180269. [CrossRef]
57. Grömping, U. Variable importance assessment in regression: Linear regression versus random forest. *Am. Stat.* **2009**, *63*, 308–319. [CrossRef]