

INSTITUTO TECNOLÓGICO DE COSTA RICA
VICERRECTORÍA DE INVESTIGACIÓN Y EXTENSIÓN
DIRECCIÓN DE PROYECTOS

PROYECTO DE INVESTIGACIÓN
INFORME FINAL

ESTIMACIÓN DE PRODUCCIÓN DE PALMA DE
ACEITE POR MEDIO DE IMÁGENES SATELITALES
DE ACCESO LIBRE, EN EL PACÍFICO CENTRAL DE
COSTA RICA

(DOCUMENTO 1)

Natalia Gómez Calderón, Ingeniería Agrícola

Fernando Watson Hernández, Ingeniería Agrícola

Cartago, marzo 2022

Tabla de contenido

1. Código y Título del Proyecto

Estimación de producción de palma de aceite por medio de imágenes satelitales de acceso libre, en el pacífico central de costa rica

2. Autores y direcciones

Dra. Natalia Gómez Calderón, ngomez@itcr.ac.cr, Escuela de Ingeniería Agrícola.

Coordinadora

Ing. Fernando Watson Hernandez, fwatson@itcr.ac.cr, Escuela de Ingeniería Agrícola

3. Resumen general del proyecto

El aceite de palma se ha convertido en uno de los aceites vegetales más consumidos en el mundo debido al elevado rendimiento sobre otros tipos de cultivos oleaginosos, además de ser un ingrediente clave en cadenas de valor globales rentables. La predicción de la edad de palma de aceite mejora con el uso de técnicas de aprendizaje automático y es una importante área de investigación en agricultura digital. El proyecto pretendió generar herramientas que permitan estimar la producción futura de palma aceitera a partir de imágenes multiespectrales satelitales de acceso libre, en plantaciones del Pacífico Central de Costa Rica. Se utilizaron técnicas de aprendizaje automático que permitieran relacionar el rendimiento de la palma aceitera con índices de vegetación entre las mismas fechas para predecir la producción a partir de futuras imágenes. Se obtuvo tres productos de la investigación, en los que se muestran las técnicas con mejores ajustes de los índices de vegetación obtenidos de las imágenes con respecto a la edad y el rendimiento de la producción, determinando que el modelo *Random Forest* para la variedad genética predominante (AVROS) es el más adecuado. También se determinó el ajuste cuando se utilizan variables para estimar escenarios de cambio climático, para el cual se estimó la necesidad de lámina de agua de riego hasta el año 2099.

4. Palabras clave

Google Earth Engine, agricultura digital, agricultura de precisión, índices de vegetación, imágenes satelitales, aprendizaje automático

5. Resultados por objetivo

Para cada uno de los objetivos, se presentan los productos obtenidos. En cada producto se incluye el resumen, palabras clave, introducción, metodología, resultados, conclusiones, referencias y agradecimientos para cada trabajo específico.

5.1. Objetivos específicos 1 y 2

Los objetivos específicos 1 y 2 son:

OE 1. Determinar las bandas espectrales que presenten mejores correlaciones con el rendimiento del cultivo de la palma aceitera para diferentes tiempos de rezago.

OE 2: Generar modelos con diferentes enfoques estadísticos y de machine learning, que permitan establecer relaciones entre los valores de longitud de onda de imágenes multispectrales y rendimiento de cultivo aplicando a diferentes rezagos de tiempo.

Sobre estos dos objetivos se escribió el artículo publicado el 03 de marzo del 2021:

Oil Palm Yield Estimation Based on Vegetation and Humidity Indices Generated from Satellite Images and Machine Learning Techniques

Referencia: Watson-Hernández, F.; Gómez-Calderón, N.; da Silva, R.P. Oil Palm Yield Estimation Based on Vegetation and Humidity Indices Generated from Satellite Images and Machine Learning Techniques. *AgriEngineering* 2022, 4, 279-291. <https://doi.org/10.3390/agriengineering4010019>

Article

Citation: Watson-Hernández, F.; Gómez-Calderón, N.; Da Silva, R. Oil Palm Yield Estimation Based on Vegetation and Humidity Indices Generated from Satellite Images and Machine Learning Techniques. *AgriEngineering* 2022, 4, x. <https://doi.org/10.3390/xxxxx>

Academic Editor: Zongmei Gao

Received: 18 January 2022

Accepted: 23 February 2022

Published: date

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



Copyright: © 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Oil Palm Yield Estimation Based on Vegetation and Humidity Indices Generated from Satellite Images and Machine Learning Techniques

Fernando Watson-Hernández ^{1,*}, Natalia Gómez-Calderón ² and Rouverson Pereira da Silva ³

^{1,2} School of Agricultural Engineering, Instituto Tecnológico de Costa Rica, 30101, Cartago, Costa Rica; fwatson@itcr.ac.cr; ngomez@itcr.ac.cr

³ Department of Engineering and Mathematical Sciences, School of Agricultural and Veterinarian Sciences, São Paulo State University (Unesp), Jaboticabal, São Paulo 14884-900, Brazil; rouverson.silva@unesp.br

* Correspondence: fwatson@itcr.ac.cr; Tel.: +50-68-4746643

Abstract: Palm oil has become one of the most consumed vegetable oils in the world, and it is a key element in profitable global value chains. In Costa Rica, oil palm cultivation is one of the three crops with the largest occupied agricultural area. The objective of this study was to explain and predict yield in safe time lags for production management by using free-access satellite images. To this end, machine learning methods were performed to a 20-year data set of an oil palm plantation located in the Central Pacific Region of Costa Rica and the corresponding vegetation indices obtained from LANDSAT satellite images. Since the best correlations corresponded to a one-year time lag, the predictive models Random Forest (RF), Least Absolute Shrinkage and Selection Operator (LASSO), Extreme Gradient Boosting (XGBoost), Recursive Partitioning and Regression Trees (RPART), and Neural Network (NN) were built for a Time-lag 1. These models were applied to all genetic material and to the predominant variety (AVROS) separately. While NN showed the best performance for multispecies information ($r^2 = 0.8139$, NSE = 0.8131, RMSE = 0.3437, MAE = 0.2605), RF showed a better fit for AVROS ($r^2 = 0.8214$, NSE = 0.8020, RMSE = 0.3452, MAE = 0.2669). The most relevant vegetation indices (NDMI, MSI) are related to water in the plant. The study also determined that data distribution must be considered for the prediction and evaluation of the oil palm yield in the area under study. The estimation methods of this study provide information on the identification of important variables (NDMI) to characterize palm oil yield. Additionally, it generates a scenario with acceptable uncertainties on the yield forecast one year in advance. This information is of direct interest to the oil palm industry.

Keywords: crop yield; google earth engine; neural network; random forest; simulation

1. Introduction

Palm oil has become one of the most consumed vegetable oils in the world due to the better extraction performance of its oil as compared to other types of oil-bearing crops such as soybeans, rapeseed or sunflower [1]. Palm oil is also a key element for profitable global value chains [2]. In Costa Rica, oil palm crops are one of the three crops with the largest occupied agricultural area. The forecasting prior to the harvest of the fresh fruit bunch is an important means to evaluate total production, which, in turn, provides useful information for the decision making related to storage, distribution, and marketing budget [3,4].

The use of empirical methods when estimating yield quantity and predicting crop production and loss leads to errors related to human factors [5]. In order to tackle the deficiencies of these methods, appropriate monthly yield forecast by means of artificial intelligence models have been created. These models describe the quantitative relation between meteorological variables with time lags and information related to the fresh fruit bunch, considering the yield of young-mature oil palm for the first six years of harvest [6].

On the other hand, being perennial trees, oil palms have a canopy structure more similar to a forest than to other agricultural crops. Therefore, remote sensing of this crop can be based on aerial or satellite images [4,6,7]. Rodríguez et al. [4] have shown that crop density estimation by sub-pixels is feasible. In agricultural applications in particular, density maps allow a more advanced analysis than that of crop land coverage binary maps. This is important for oil palm plantations, where the distance between trees is known to correlate with production yield.

To overcome the limitations of the monitoring images obtained, multiple spectral bands were used in modeling the colors, estimating vegetation indices (VIs), or in doing a spectral mixture analysis (SMA) [7]. Observational data based on spectral reflectance has been widely used to monitor crop conditions and estimate their yield [8]. The bank of information per pixel and the information from the crops allow the creation of correlations from algorithms that facilitate the analysis of the data. Khanal et al. [9] recommend evaluating various machine learning algorithms to improve the precision of the prediction estimates as well as the evaluation of soil parameters and crops yield. This is done by using data division relations for the training and validation of the model with a 7:3 or 8:2 ratio [10–13]. Amirruddin et al. [13] state that using machine learning techniques to manage data is a promising field for the evaluation of oil palm crops since it reduces costs, time, and intensive work for the wide plantation areas.

Since the LANDSAT satellite was launched into orbit in 1972, free-access multispectral satellite images are available for the region. The LANDSAT images provide information of up to 11 wavelengths (bands), with resolutions of between 15 m and 100 m, which are acceptable ranges due to the fact that the management areas for oil palm cultivation in Costa Rica exceed four hectares [14]. Due to the climatic variability of the Central Pacific region of Costa Rica, it is important to find relationships between images multispectrality, yield, climatic and edaphic conditions, and the age of the crop. This allows explaining and predicting production in safe time lags for the management of production.

The objective of this work was to generate a tool that facilitates the prediction of crop yield. To this end, several statistical models were designed to estimate the annual production of the oil palm crop by means of the vegetation and humidity indices generated from the Landsat 5, 7 and 8 images in the Central Pacific of Costa Rica and the time lag with better correlation between index and crop yield.

2. Materials and Methods

The study was carried out with the information of an oil palm crop from an agribusiness located in the Central Pacific of Costa Rica (Figure 1). The crop has a 9×9 m field spacing between plants. The analyzed plantation consisted mainly of the AVROS variety (40%), EKONA (21%), and others in the process of being replaced by AVROS (39%). One hundred and three productive units (103 PU) were considered, with information of between 16 and 20 years of cultivation on genetic variety, crop yield (t/ha) and year of sowing. Such variables were averaged monthly for each year and the best correlation was identified for time lags of one, two, three or four years.

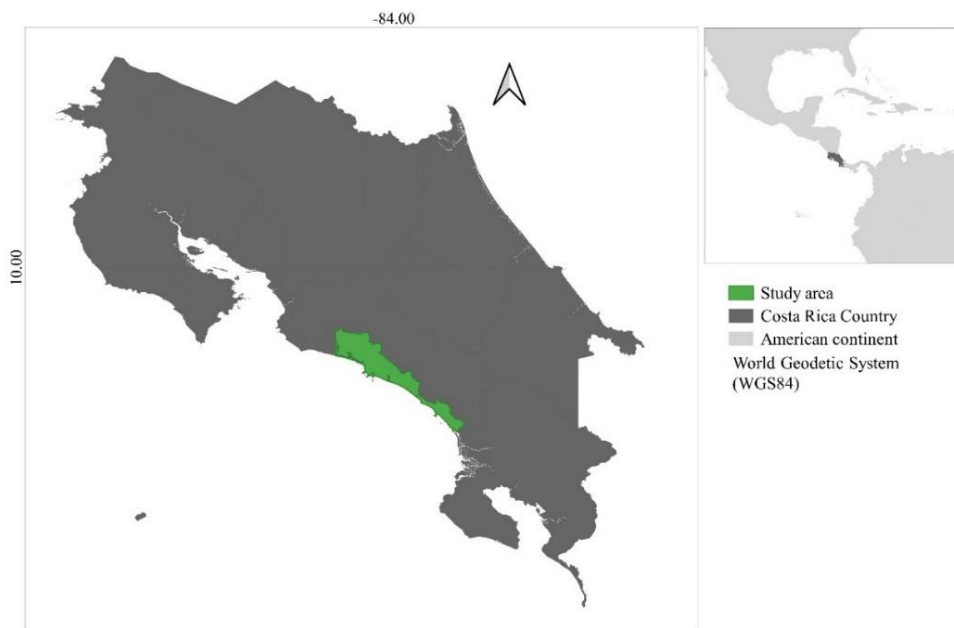


Figure 1. Location of the oil palm plantation under study.

For the productive units considered, the images Landsat 5, 7 and 8 Collection 1 Tier 1 calibrated top-of-atmosphere (TOA) reflectance [15] corresponding to a period between January 1996 and December 2016 were obtained, in the ranges described in Table 1. Data management was done through a Java Script code on Google Earth Engine (GEE). The code cyclically generates 12 indices of both vegetation and humidity for each image of the collection (Table 2), allowing to correlate the spectral information of the images with the biophysical properties of the vegetation cover [16]. This generates a new 20 year-history collection with an inter-annual resolution for each of the indices used. As input, the tool requires

information of the geometry of each one of the 103 PUs, which constitute the minimum information unit used about production. From the layers of indices and the delimited areas, a statistical description of the behavior of the index is generated in each PU for each date of image capture. The information consists of minimum and maximum values, quartiles 25th and 75th, means, medians and standard deviations. The information is ordered chronologically in a file for each PU with a total of 84 variables. For the selection of the photographs used, there was no discrimination within the range of dates established by any parameter, so that the models generated were in charge of assigning greater or lesser weights to each variable when using the values of maximum, minimum, percentiles and standard deviation of the vegetation index within the PUs.

Table 1. The standard spectral bands of the optical range of the images used in this study.

Bands	Wavelength (μm)		
	Landsat 5	Landsat7	Landsat8
1 (BLUE)	0.45–0.52	0.441–0.514	
2 (BLUE)			0.442–0.5120
2 (GREEN)	0.52–0.60	0.519–0.601	
3 (GREEN)			0.533–0.590
3 (RED)	0.63–0.69	0.631–0.692	
4 (RED)			0.636–0.673
4 (NIR)	0.76–0.90	0.772–0.898	
5 (NIR)			0.851–0.879
5 (SWIR)	1.55–1.75	1.547–1.749	
6 (SWIR)			1.566–1.651

Table 2. Multispectral indices from spectral channels using the Landsat 5, Landsat 7 and Landsat 8 collections.

Variable	Index	Equation	Source
Vegetation	ARVI	$(NIR - (2 \cdot RED) + BLUE)/(NIR + (2 \cdot RED) + BLUE)$	[17]
Vegetation	AVI	$[NIR \cdot (1 - RED) \cdot (NIR - RED)]^{1/3}$	[18]
Vegetation	EVI	$2.5 \cdot (NIR - RED)/(NIR + 6 \cdot RED - 7.5 \cdot BLUE + 1)$	[19]
Vegetation	GCI	$(NIR/GREEN) - 1$	[17]
Vegetation	GNDVI	$(NIR - GREEN)/(NIR + GREEN)$	[19]
Vegetation	NDVI	$(NIR - RED)/(NIR + RED)$	[19]
Vegetation	NPCRI	$(RED - BLUE)/(RED + BLUE)$	[18]
Vegetation	SAVI	$1.5 \cdot (NIR - RED)/(NIR + RED + 0.5)$	[19]
Vegetation	SIPI	$(NIR - BLUE)/(NIR + BLUE)$	[19]
Water	MSI	$(SWIR/NIR)$	[20]
Water	NDMI	$(NIR - SWIR)/(NIR + SWIR)$	[20]
Water	NDWI	$(GREEN - NIR)/(GREEN + NIR)$	[20]

RStudio (2018) was used to merge information about the crops yield with the vegetation and humidity indices. The process consisted of changing the time scale from monthly data to annual averages, eliminating the missing data, and overlapping the data of indices with time lags of one,

two, three and four years (time lags = 1, 2, 3 and 4 years) with respect to the series of performance data.

In selecting the time lag to work with, the individual correlation was determined for each index considering the performance of each of the proposed time lags. The time lag with the best correlation was used to build the AI models. The models used were Random Forest (RF), XGBoost algorithm (XGBoost), LASSO regression (LASSO), Recursive Partitioning and Regression Trees (RPART), and Neural Network (NN); the packages used were randomForest, xgb.train, glmnet, rpart and neuralnet, respectively, to which subroutines were generated in RStudio. The configuration settings of the proposed models are in Table 3.

Table 3. Parameters used in the configuration of the proposed models.

Model	Parameters	Description	Source
LASSO	alpha	The elasticnet mixing parameter, with $0 \leq \alpha \leq 1$	[21]
	lambda	Regularization hyperparameter	
RF	ntree	Number of trees to grow	[22]
	mtry	Number of variables randomly sampled as candidates at each split	
XGBoost	max.depth	Maximum depth of a tree	[23]
	nrounds	The number of decision trees in the final model	
	nthread	Number of parallel threads used to run XGBoost	
	objective	Specify the learning task and the corresponding learning objective	
RPART	minsplit	The minimum number of observations that must exist in a node in order for a split to be attempted.	[24]
	minbucket	The minimum number of observations in any terminal node	
	cp	Complexity parameter	
NN	threshold	A numeric value specifying the threshold for the partial derivatives of the error function as stopping criteria	[25]
	stepmax	The maximum steps for the training of the neural network	
	algorithm	A string containing the algorithm type to calculate the neural network	

As seen in Figure 2, the dependent variable was the performance with a time lag of one year and the predictor variables were the vegetation and humidity indices. The set was then randomly divided into the training (Data Training) and validation (Data Test) stages in a 7:3 ratio. The Data Test was further subdivided into two series of information. The first consisted of data from all varieties of oil palm (genetic material) produced in the time interval under study, however, the second consisted of the information from the AVROS palm variety, given that it is the predominant genetic material.

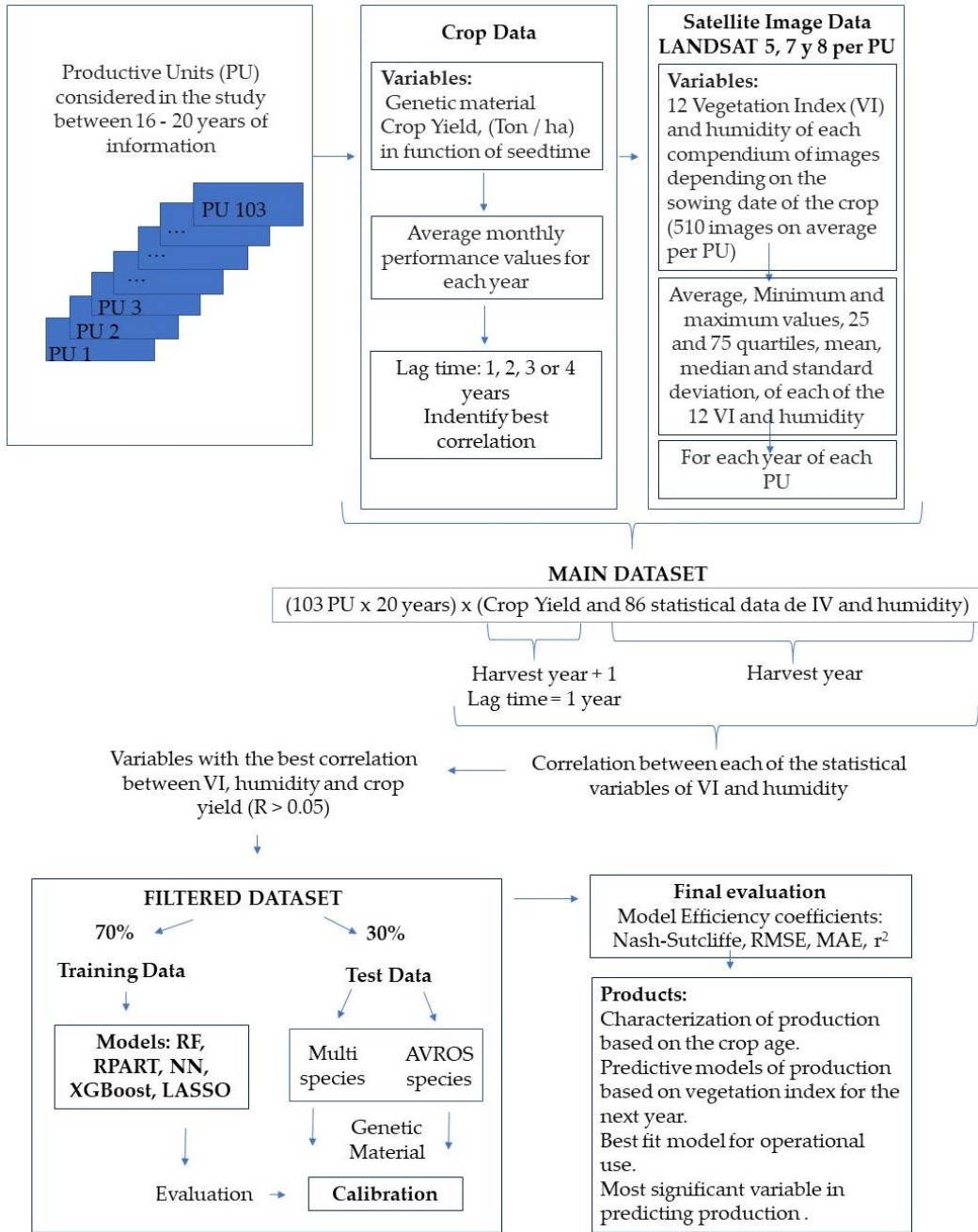


Figure 2. Methodological diagram of the analysis of the information and model's sensitivity.

The evaluation of the models was carried out by applying the efficiency coefficient of the Nash–Sutcliffe model (NSE) (Equation (1)), the root mean square error (RMSE) (Equation (2)), the mean absolute error coefficient (MAE) (Equation (3)), and the coefficient of determination (r^2) (Equation (4)).

$$NSE = 1 - \frac{\sum_{i=1}^n (y_i^{obs} - y_i^{sim})^2}{\sum_{i=1}^n (y_i^{obs} - \bar{y}^{obs})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i^{obs} - y_i^{sim})^2} \quad (2)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i^{obs} - y_i^{sim}| \quad (3)$$

$$r^2 = \left(\frac{\sum [(y_i^{obs} - \bar{y}^{obs})(y_i^{sim} - \bar{y}^{sim})]}{\sqrt{\sum (y_i^{obs} - \bar{y}^{obs})^2 * \sum (y_i^{sim} - \bar{y}^{sim})^2}} \right)^2 \quad (4)$$

where n is the amount of data available for the study, y_i^{obs} corresponds to the information on the crop yield obtained for the interval under study, \bar{y}^{obs} is the average of the crop yield obtained for the interval under study, y_i^{sim} is the simulated crop yield, and \bar{y}^{sim} is the average simulated crop yield.

The best model for each of the data series used in the validation process (AVROS and Multi Species) was determined by comparing the values of the efficiency and error coefficients obtained for each model (RF, LASSO, XGBoost, RPART, NN). For the best qualified model according to the data series of the data test, the importance of the vegetation and humidity indices was determined, following the procedure described below.

-Variable importance Random Forest Model

In determining the importance of the variables for the Random Forest model, the Mean Decrease Gini ($I_G(\theta)$) method was used, as shown in Equation (5). The Mean Decrease Gini is a measure that dimensions the importance of the variables based on the Gini impurity index ($i(\tau)$) used to calculate the divisions of the trees. A loss function (mse) is used where, by comparison, it is established that the most useful variables achieve greater increases in the purity of the nodes [26].

$$I_G(\theta) = \sum_T \sum_\tau \Delta i_\theta(\tau, T) \quad (5)$$

where, considering all the variables (θ) used in each node (τ) within the trees (T) of the Random Forest model, the number of optimal divisions is established, from which the decrease in the Gini impurity $\Delta i_\theta(\tau, T)$ is calculated, which is registered and accumulated for each node τ of each tree (T) of the Random Forest model individually for each of the variables (θ).

-Variable importance Neural Network model

To determine the importance of the variables of the Neural Network model, Garson's algorithm method [27] was used. This identifies the relative importance of the independent variables in a NN by deconstructing the weights assigned by the model, according to the Equation (6). The relative importance of an independent variable in response to the dependent variable is identified by locating all connections through the weighting of the nodes of interest. This procedure is repeated for all the independent variables used in the model [28].

$$CR_{ik} = \frac{\sum_{j=1}^L \left(\frac{w_{ij}}{\sum_{r=1}^N w_{rj}} v_{jk} \right)}{\sum_{i=1}^N \left(\sum_{j=1}^L \left(\frac{w_{ij}}{\sum_{r=1}^N w_{rj}} v_{jk} \right) \right)} \quad (6)$$

where, CR_{ik} represents the percentage of influence of each input independent variable i on the output dependent variable $\sum_{r=1}^N w_{rj}$, being the sum of the weights that connect the input layer i and the neuron j ; N corresponds to the total of input variables; L corresponds to the total of the hidden layer; v_{jk} that corresponds to the weights of the connection between the input neuron j and the input vector k .

The information on the yield of the oil palm fruit production used has a normal behavior (Figure 3).

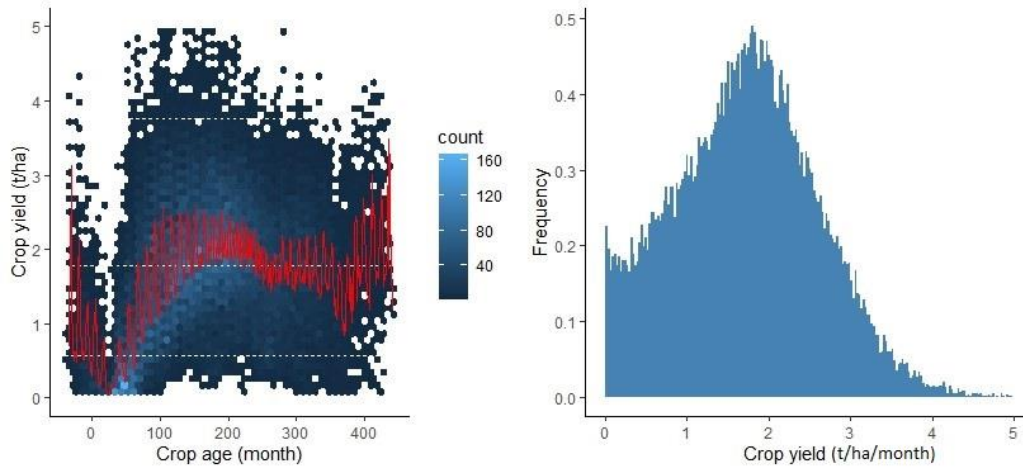


Figure 3. Characterization of the production behavior of oil palm.

3. Results

The analysis of the yield of the 103 PUs used in the creation of the models determined that, between month 0 and 84, there is a growth in yield of 0.26 t/ha per year. Thereafter, a constant production is relatively maintained up to approximately year 20 of the crop, where the most frequent production value is 1.92 t/ha per month (Figure 3).

Correlations between vegetation and humidity indices for different time lags

Figure 4 shows the value of the correlation coefficient for Time-lags of one, two, three and four years. The 15 vegetation indices with the highest absolute correlations with crop yield were selected. Positive values indicate that the higher the index value, the better the plantation yield and, on the contrary, indices that show water stress (MSI) [29] or the identification of senescence in the plantation (NPCRI) [30] will have a inversely proportional relationship.

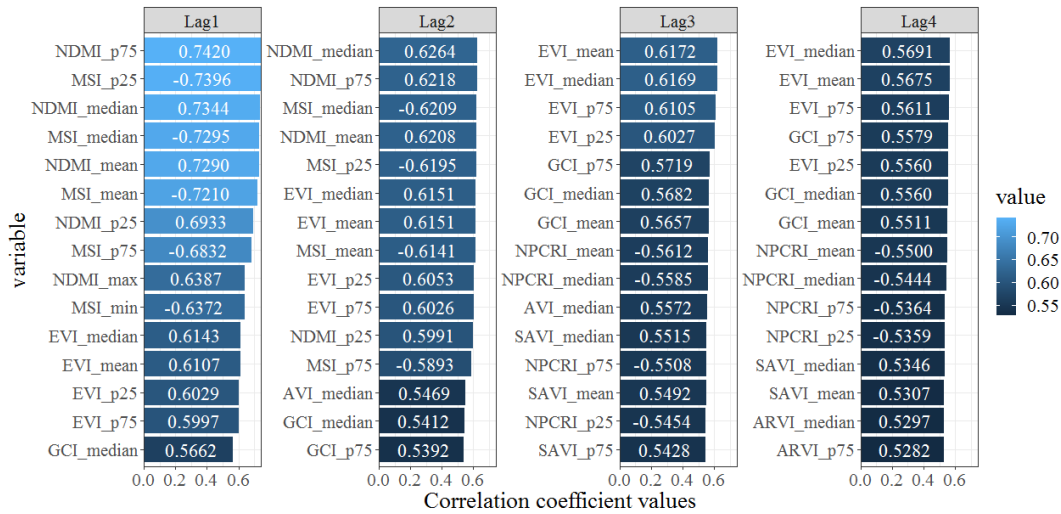


Figure 4. Crop yield correlation coefficients with the variables generated for time lags of one, two, three and four years.

Moisture-related indices (NDMI, MSI), either by amount of water or by plant stress, show the highest correlations with yield for lags of one and two years (Lag-time 1 and Lag-time 2). After the three-year delay time (Time Lag- 3), humidity is not significant in the analysis, and the vegetation indices relative to chlorophyllous activity (EVI, GCI, NPCRI) show the best yield relations.

The correlation values of the indices relative to chlorophyllous activity does not show variation in any of the delay times studied (Figure 4). In any case, the highest value reached is $r = 0.63$ for a lag of three years.

The best correlations were obtained for the Time-Lag of one year, as its first six variables present values above $r = 0.70$, while the lags from two to four years do not exceed $r = 0.63$. Therefore, the prediction models were built for a Time-lag 1.

Coefficients of crop production prediction models

Machine learning models were built according to the RF, LASSO, XGBoost, RPART and NN structures. Table 4 shows the results obtained in the calibration of the models. For the Values NSE and r^2 , acceptability ranges according to [31] were followed; whereas for RMSE and MAE, the average production yield (1.92 t/ha) was used. A 10% error was considered acceptable, and more than 40% unsatisfactory. For the RF model, three different values according to the number of trees (ntree) were tested: 5000, 15,000 and 10,000. The best performance according to the statistical evaluation coefficients used was obtained for ntree = 10,000.

Table 4. Calibration evaluation coefficients for the models used.

Model	Variable /Value	NSE	RMSE	MAE
LASSO	lambda= 0.0037	0.7701	0.3487	0.2679
Neural Network (NN)	stepmax = 1e+05	Algorithm did not converge		
	stepmax = 1e+06	Algorithm did not converge		
	stepmax = 1e+07	0.7947	0.3925	0.2514
Random Forest (RF)	ntree = 5,000	0.9519	0.1668	0.1282
	ntree = 10,000	0.9527	0.1655	0.1266
	ntree = 15,000	0.9518	0.167	0.1279

Recursive Partitioning and Regression Trees (RPART)	Minsplit =	5	0.6652	0.4207	0.3217
		10	0.6652	0.4207	0.3218
		50	0.6652	0.4207	0.3218
		100	0.6652	0.4207	0.3217
		500	0.5495	0.488	0.3765
		1000	0.4133	0.557	0.4364
	Complexity parameter (cp) = minsplit = 10	0.01	0.6652	0.4207	0.3218
		0.001	0.8875	0.2438	0.1876
		0.0001	0.9322	0.1894	0.1324
	XGBoost	max.depth = 1	0.6777	0.4127	0.3180
max.depth = 2		0.8230	0.3059	0.2322	
max.depth = 3		0.9034	0.2259	0.1727	
max.depth = 4		0.9567	0.1512	0.1147	
max.depth = 5			over fitting		

The LASSO linear regression method (glmnet package) was included in the analysis as a reference method. The lambda that generated the lowest MSE was used to calibrate it ($\lambda = 0.0037$).

For RPART, the minsplit parameter was used first to improve performance. The minsplit values 5, 10, 50, 100, 500 and 1000 were tested, and Minsplit = 10 was selected. The model fit continued with the variable complexity parameter (cp) for values of 0.01, 0.001 and 0.0001, being the latter the one used. In the NN model, a stepmax of 1×10^7 was used as it presented convergence problems for lower values (Figure 4).

Calibration (training) and validation (test) of prediction models for crop production

As seen in Figure 5, the five models generated have a satisfactory performance for the training, with values of $r^2 > 0.7900$, $NSE > 0.7900$ and $RMSE \leq 0.3300$ t/ha. The best model in the training was XGBoost with $r^2 = 0.9597$, $NSE = 0.9567$ and $RMSE = 0.1513$ t/ha, followed by RF with a slightly lower performance. RPART presented atypical residue values higher than XGBoost and RF (Figure 6), which generates an increase in $RMSE = 0.1894$ t/ha. Nevertheless, these atypical data in the residues are present in yields between 1–2 t/ha, where the model tends to overestimate palm production in some dispersed points. Its performance, however, improves for values higher than 2 t/ha. Since the average production of the time interval of the crop under study is 1.92 t/ha, the RPART model can be used with a 98% confidence for estimated values greater than the average. The LASSO and NN models presented a variation in the residues greater than RF, XGBoost and RPART, which generated an $RMSE$ higher than 0.3200 t/ha, 59% more than RF, XGBoost and RPART. The LASSO reference model obtained acceptable coefficients of $NSE = 0.7701$, $r^2 = 0.7702$ during the training process, however the performance was slightly lower than the average of the other machine learning methods. The LASSO model generated coefficient values of $RMSE = 0.3487$ t/ha, $MAE = 0.2679$ t/ha that are within the range considered acceptable, yet these errors are higher than the average errors of the best performing machine learning models (RF, XGBoost).

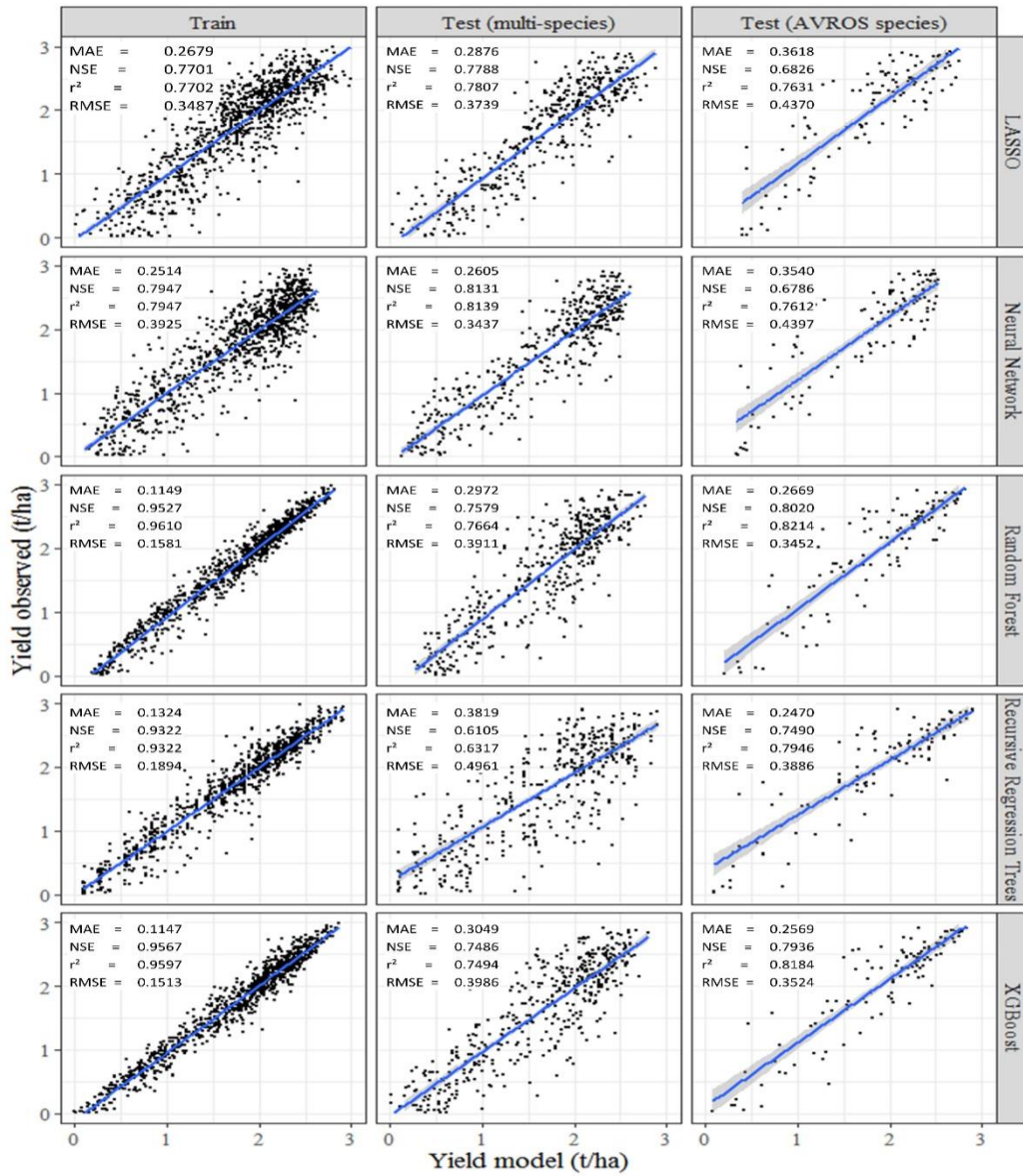


Figure 5. Comparison of observed performance with simulated performance in the calibration and validation processes.

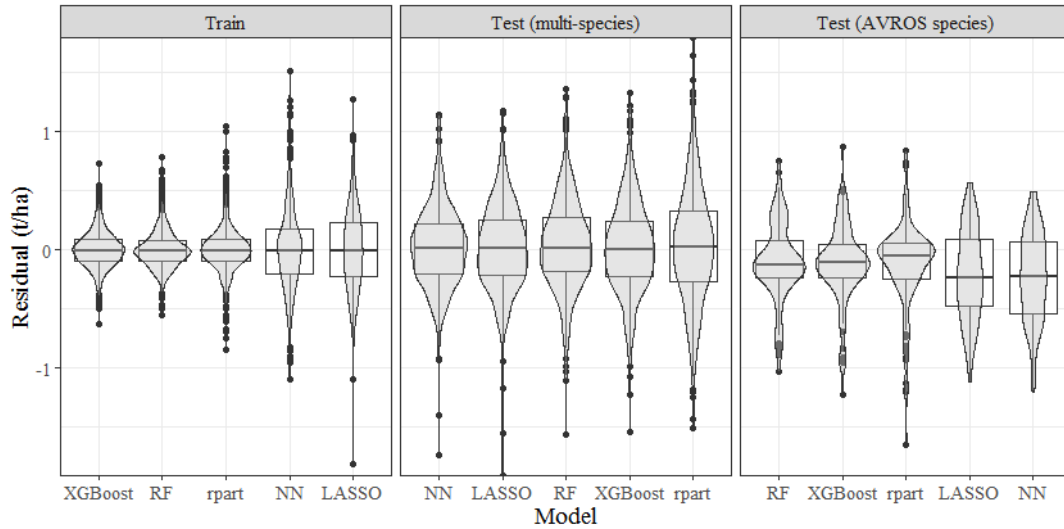


Figure 6. Residual distribution of the models in the calibration and validation processes. The dots indicate the value of the RMSE.

For validation the data set consisted of information from multiple palm species including all the genetic materials used in the cultivation between 1996 and 2016 (Multi-species Test). In this stage, the evaluation coefficients of the models decreased, presenting values of r^2 between 0.6400–0.8300, NSE between 0.6100–0.8100, and the RMSE between 0.3400–0.5000 t/ha. In descending order of the evaluation coefficients, the best model was NN, followed by LASSO, RF, XGBoost and RPART. RPART presented a higher error than (RMSE = 0.4961 t/ha) the other models, thus, it is the least recommended model to estimate yield from information that considers multiple species (genetic material).

For the validation based on the information of the crop yield of the AVROS species (24.9% of the total data), the models behaved similarly to the multi-species test, presenting r^2 values between 0.7600–0.8100, NSE between 0.6800–0.8000 and RMSE between 0.3400–0.4400 t/ha. However, following the same ordering criterion of the calibration models, the best model was RF, followed by XGBoost, RPART, LASSO and NN. Consequently, NN is the least recommended model when using only the information of the predominant genetic material (AVROS). The selection of the model is influenced by the variability of the genetic material used in planting, as this alters the performance of the machine learning techniques.

Importance of vegetation indices variables in crop prediction

Based on the evaluation coefficients of the models in both training and validation together with residue behavior, it is determined that the NN model has the best performance for the multispecies Test, while for the AVROS species test, the RF model shows the best fit. When performing the analysis of the variables of the most important vegetation indices (VIs) to predict from these two models (Figure 7), the median of the NDMI (NDMI_median) stands out as the most important variable for both models.

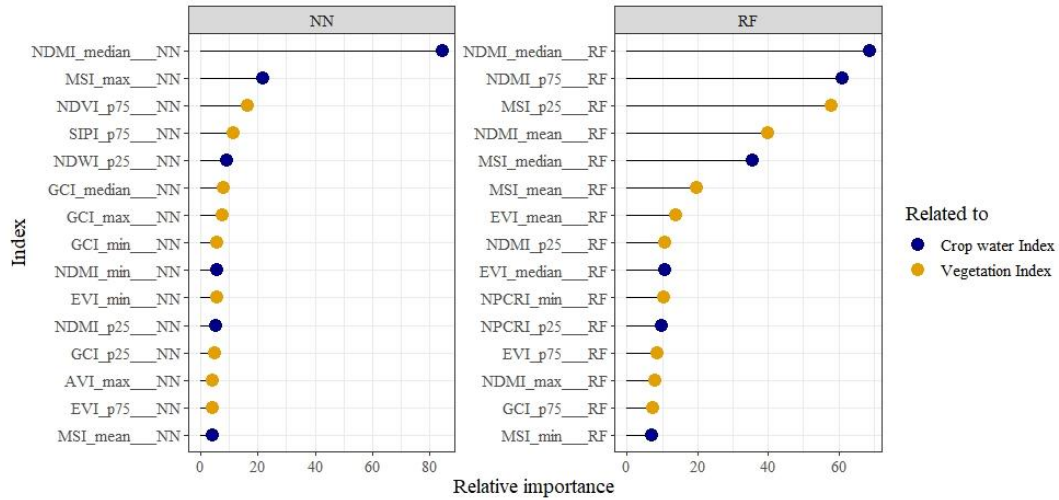


Figure 7. Coefficients of the relative importance of the variables used in the two best models.

4. Discussion

The most important VIs obtained from the Landsat 5, 7 and 8 satellite images are related to water in the plant. The 75th and 25th percentiles of these VIs improve the performance of the models compared to other studies where they were not considered (Table 5). In the two best models NN (multispecies Test) y RF (AVROS species test) of the fifteen most important variables, six correspond to the percentiles of the VIs (25th and 75th quartiles); that is, the distribution of the available information on the VIs of each PU impacts the performance of the models significantly and they should be considered together with the mean values of the VIs in predicting and evaluating oil palm production.

Table 5. Previous research works on oil palm crop yield prediction.

Method	Variables	Coefficient	Source
Artificial Neural Network	NDVI	$r^2 = 0.5100$	[32]
Genetic algorithm	Historical yield data, cropland information, climatic information, air pollutants	$r^2 = 0.9400$ RMSE = 0.1500 t/ha	[33]
Artificial Neural Network	Climatic information	MAE = 0.5300 t/ha RMSE = 0.6800 t/ha	[34]
Long short-term memory (LSTM)	Historical yield data	MAPE = 2.7100%	[35]

The study by Hilal et al. [33] obtained a better performance than in this study (the difference in RMSE is 0.1950t/ha); however, the data of the model proposed in this study were obtained from satellite images and did not require additional data measured on site, which reduces its difficulty of application.

In the prediction of the performance that considered the totality of the genetic material, the distribution of the residues in the validation stage shows an average value near zero (mean = 0.0300 t/ha) and a homogeneous distribution with a higher concentration between quartiles 25th and 75th, for the NN, RF and XGBoost models. On the other hand, when considering

only the variety AVROS, the mean of the residues during validation is less than zero in all the models. Thus, the distribution of the residues for AVROS species test losses homogeneity compared to the Multi Specie Test, which will cause the crop yield to be underestimated in the prediction. Due to this, the NN model for the Multi-specie Test would not differentiate the significance of some variables.

In predicting the production behavior of the oil palm crop with any genetic material by means of machine learning techniques (RF, RPART, LASSO, XGBoost and NN) and vegetation indices variables obtained from Landsat 6, 7 and 8 satellite images, the appropriate time lag is 1 year. The NN model is the one with the best performance for crop information considering multiple genetic planting materials. However, NN is the least recommended model when using only the predominant genetic material information. When considering the predominant genetic material (AVROS), the RF model shows the best performance. On the other hand, RPART presented a higher error than (RMSE = 0.4961 t/ha) the other models, and thus, it is the least recommended to estimate yield from information that considers multiple species.

Our results show that the variability of the genetic material used in planting influences the selection of the model. This is due to performances of machine learning techniques differing based on genetic variability. The vegetation indices related to water in the plant [36] and the data, corresponding to the average and the 25th and 75th percentiles, are the variables that have the greatest influence on the performance of the proposed prediction models. In the best models (NN and RF), six of the fifteen most important variables correspond to percentile vegetation indices (25th and 75th quartiles). This shows that contemplating the distribution of the data within the PUs significantly improves the performance of the models. As a result of this, they must be considered in the prediction and evaluation of oil palm production in the study area.

5. Conclusions

The strongest correlations between oil palm yield and VIs were obtained for a lag period of one year. The machine learning methods used to estimate oil palm crop yields as a function of VIs one year in advance showed satisfactory performance. The RF model was the best qualified for predicting oil palm cultivation of the AVROS species (MAE = 0.2669t/ha, RMSE = 0.3452t/ha, NSE = 0.8020, $r^2 = 0.8214$), while the NN model was the best when the plantation has multiple species (MAE = 0.2605, RMSE = 0.3437, NSE = 0.8131, $r^2 = 0.8139$). The Normalized Difference Moisture Index (NDMI) is the most relevant variable in the prediction of oil palm cultivation among a total of 12 VIs used, regardless of the type of species under study [37]. The estimation methods of this study can provide information on the identification variables (NDMI) to characterize palm oil yield. Furthermore, it generates a scenario with acceptable uncertainties on the yield forecast one year in advance, which is of direct interest to the palm oil industry.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Figure S1: title; Table S1: title; Video S1: title.

Author Contributions: Conceptualization, F.W.-H. and N.G.-C.; methodology, F.W.-H.; software, F.W.-H.; validation, F.W.-H. and N.G.-C.; formal analysis, F.W.-H., N.G.-C. and R.P.D.S.; investigation, F.W.-H. and N.G.-C.; resources, F.W.-H.; data curation, F.W.-H.; writing—original draft preparation, F.W.-H. and N.G.-C.; writing—review and editing, F.W.-H., N.G.-C. and R.P.D.S.; visualization, N.G.-C.; supervision, F.W.-H., N.G.-C.; project administration, N.G.-C.; funding acquisition, F.W.-H. All authors have read and agreed to the published version of the manuscript." Please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement:

Acknowledgments: We thank the oil palm producers of the Central Pacific of Costa Rica for providing the historical records of production yields and the Vice-Rector's Office for Research and Extension of the Technological Institute of Costa Rica for supporting the study.

Conflicts of Interest: Declare conflicts of interest or state "The authors declare no conflict of interest." Authors must identify and declare any personal circumstances or interest that may be perceived as inappropriately influencing the representation or interpretation of reported research results. Any role of the funders 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 must be declared in this section. If there is no role, please state "The funders 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. Camacho, A.; Correa, C.V.; Arguello, H. An Analysis of Spectral Variability in Hyperspectral Imagery: A Case Study of Stressed Oil Palm Detection in Colombia. *Int. J. Remote Sens.* **2019**, *40*, 7603–7623. <https://doi.org/10.1080/01431161.2019.1595210>.
2. Castellanos-Navarrete, A.; de Castro, F.; Pacheco, P. The Impact of Oil Palm on Rural Livelihoods and Tropical Forest Landscapes in Latin America. *J. Rural Stud.* **2021**, *81*, 294–304. <https://doi.org/10.1016/j.jrurstud.2020.10.047>.
3. da Silva Cabral de Moraes, J.R.; Souza Rolim, G.D.; Martorano, L.G.; de Oliveira Aparecido, L.E.; Padilha de Oliveira, M.D.S.; de Farias Neto, J.T. Agrometeorological Models to Forecast Açai (*Euterpe Oleracea* Mart.) Yield in the Eastern Amazon. *J. Sci. Food Agric.* **2020**, *100*, 1558–1569. <https://doi.org/10.1002/jsfa.10164>.
4. Rodríguez, A.; Ruiz-Ramos, M.; Palosuo, T.; Carter, T.R.; Fronzek, S.; Lorite, I.J.; Ferrise, R.; Pirttioja, N.; Bindi, M.; Baranowski, P.; et al. Implications of Crop Model Ensemble Size and Composition for Estimates of Adaptation Effects and Agreement of Recommendations. *Agric. For. Meteorol.* **2019**, *264*, 351–362. <https://doi.org/10.1016/j.agrformet.2018.09.018>.
5. Akhter, R.; Sofi, S.A. Precision Agriculture Using IoT Data Analytics and Machine Learning. *J. King Saud Univ.-Comput. Inf. Sci.* **2021**, in press. <https://doi.org/10.1016/j.jksuci.2021.05.013>.
6. Rodríguez, A.C.; D'Aronco, S.; Schindler, K.; Wegner, J.D. Mapping Oil Palm Density at Country Scale: An Active Learning Approach. *Remote Sens. Environ.* **2021**, *261*, 112479. <https://doi.org/10.1016/j.rse.2021.112479>.
7. Morel, A.C.; Fisher, J.B.; Malhi, Y. Evaluating the Potential to Monitor Aboveground Biomass in Forest and Oil Palm in Sabah, Malaysia, for 2000–2008 with Landsat ETM+ and ALOS-PALSAR. *Int. J. Remote Sens.* **2012**, *33*, 3614–3639. <https://doi.org/10.1080/01431161.2011.631949>.

8. Ryu, J.-H.; Oh, D.; Cho, J. Simple Method for Extracting the Seasonal Signals of Photochemical Reflectance Index and Normalized Difference Vegetation Index Measured Using a Spectral Reflectance Sensor. *J. Integr. Agric.* **2021**, *20*, 1969–1986. [https://doi.org/10.1016/S2095-3119\(20\)63410-4](https://doi.org/10.1016/S2095-3119(20)63410-4).
9. Khanal, S.; Fulton, J.; Klopfenstein, A.; Douridas, N.; Shearer, S. Integration of High Resolution Remotely Sensed Data and Machine Learning Techniques for Spatial Prediction of Soil Properties and Corn Yield. *Comput. Electron. Agric.* **2018**, *153*, 213–225. <https://doi.org/10.1016/j.compag.2018.07.016>.
10. Suharjito; Elwirehardja, G.N.; Prayoga, J.S. Oil Palm Fresh Fruit Bunch Ripeness Classification on Mobile Devices Using Deep Learning Approaches. *Comput. Electron. Agric.* **2021**, *188*, 106359. <https://doi.org/10.1016/j.compag.2021.106359>.
11. Mamehgoel Yousefi D.B.; Mohd Rafie, A.S.; Abd Aziz, S.; Azrad, S.; Mazmira Mohd Masri, M.; Shahi, A.; Marzuki, O.F. Classification of Oil Palm Female Inflorescences Anthesis Stages Using Machine Learning Approaches. *Inf. Process. Agric.* **2020**, *in press*. <https://doi.org/10.1016/j.inpa.2020.11.007>.
12. Zheng, J.; Fu, H.; Li, W.; Wu, W.; Yu, L.; Yuan, S.; Tao, W.Y.W.; Pang, T.K.; Kanniah, K.D. Growing Status Observation for Oil Palm Trees Using Unmanned Aerial Vehicle (UAV) Images. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 95–121. <https://doi.org/10.1016/j.isprsjprs.2021.01.008>.
13. Amirruddin, A.D.; Muharam, F.M.; Ismail, M.H.; Tan, N.P.; Ismail, M.F. Hyperspectral Spectroscopy and Imbalance Data Approaches for Classification of Oil Palm's Macronutrients Observed from Frond 9 and 17. *Comput. Electron. Agric.* **2020**, *178*, 105768. <https://doi.org/10.1016/j.compag.2020.105768>.
14. Instituto Nacional de Estadística y Censos [INEC]. *VI Censo Nacional Agropecuario: Características de Las Fincas y de Las Personas Productoras*, 1st ed.; Instituto Nacional de Estadística y Censos [INEC]: San José, Costa Rica, 2015; ISBN 978-9968-683-99-9.
15. Vos, K.; Harley, M.D.; Splinter, K.D.; Simmons, J.A.; Turner, I.L. Sub-Annual to Multi-Decadal Shoreline Variability from Publicly Available Satellite Imagery. *Coast. Eng.* **2019**, *150*, 160–174. <https://doi.org/10.1016/j.coastaleng.2019.04.004>.
16. Aljahdali, M.O.; Munawar, S.; Khan, W.R. Monitoring Mangrove Forest Degradation and Regeneration: Landsat Time Series Analysis of Moisture and Vegetation Indices at Rabigh Lagoon, Red Sea. *Forests* **2021**, *12*, 52. <https://doi.org/10.3390/f12010052>.
17. Dube, T.; Pandit, S.; Shoko, C.; Ramoelo, A.; Mazvimavi, D.; Dalu, T. Numerical Assessments of Leaf Area Index in Tropical Savanna Rangelands, South Africa Using Landsat 8 OLI Derived Metrics and In-Situ Measurements. *Remote Sens.* **2019**, *11*, 829. <https://doi.org/10.3390/rs11070829>.
18. Sykas, D. Spectral Indices with Multispectral Satellite Data. Available online: <https://www.geo.university/pages/blog?p=spectral-indices-with-multispectral-satellite-data> (accessed on 24 November 2021).
19. Tan, C.; Zhou, X.; Zhang, P.; Wang, Z.; Wang, D.; Guo, W.; Yun, F. Predicting Grain Protein Content of Field-Grown Winter Wheat with Satellite Images and Partial Least Square Algorithm. *PLoS ONE* **2020**, *15*, e0228500. <https://doi.org/10.1371/journal.pone.0228500>.
20. Zakir, M. GIS and RS Based Spatio-Temporal Analysis of Soil Moisture/Water Content Variation in Southern Irrigated Part of Sindh, Pakistan. *J. Remote Sens. GIS* **2018**, *7*, 250. <https://doi.org/10.4172/2469-4134.1000250>.
21. Li, Z.; Sillanpää, M.J. Overview of LASSO-Related Penalized Regression Methods for Quantitative Trait Mapping and Genomic Selection. *Theor. Appl. Genet.* **2012**, *125*, 419–435. <https://doi.org/10.1007/s00122-012-1892-9>.
22. Huo, J.; Shi, T.; Chang, J. Comparison of Random Forest and SVM for Electrical Short-Term Load Forecast with Different Data Sources. In Proceedings of the 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 26–28 August 2016; IEEE: New York, NY, USA, 2017; pp. 1077–1080. <https://doi.org/10.1109/ICSESS.2016.7883252>.
23. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794. <https://doi.org/10.1145/2939672.2939785>.
24. Nembrini, S. Prediction or Interpretability? *Emerg. Themes Epidemiol.* **2019**, *16*, 4. <https://doi.org/10.1186/s12982-019-0086-1>.
25. Günther, F.; Fritsch, S. Neuralnet: Training of Neural Networks. *R J.* **2010**, *2*, 30–38.
26. Menze, B.H.; Kelm, B.M.; Masuch, R.; Himmelreich, U.; Bachert, P.; Petrich, W.; Hamprecht, F.A. A Comparison of Random Forest and Its Gini Importance with Standard Chemometric Methods for the Feature Selection and Classification of Spectral Data. *BMC Bioinform.* **2009**, *10*, 213. <https://doi.org/10.1186/1471-2105-10-213>.

27. da Costa, N.L.; de Lima, M.D.; Barbosa, R. Evaluation of Feature Selection Methods Based on Artificial Neural Network Weights. *Expert Syst. Appl.* **2021**, *168*, 114312. <https://doi.org/10.1016/j.eswa.2020.114312>.
28. Li, P.; Hua, P.; Gui, D.; Niu, J.; Pei, P.; Zhang, J.; Krebs, P. A Comparative Analysis of Artificial Neural Networks and Wavelet Hybrid Approaches to Long-Term Toxic Heavy Metal Prediction. *Sci. Rep.* **2020**, *10*, 13439. <https://doi.org/10.1038/s41598-020-70438-8>.
29. Watt, M.S.; Leonardo, E.M.C.; Estarija, H.J.C.; Massam, P.; de Silva, D.; O'Neill, R.; Lane, D.; McDougal, R.; Buddenbaum, H.; Zarco-Tejada, P.J. Long-Term Effects of Water Stress on Hyperspectral Remote Sensing Indicators in Young Radiata Pine. *For. Ecol. Manag.* **2021**, *502*, 119707. <https://doi.org/10.1016/j.foreco.2021.119707>.
30. Sosa, L.; Justel, A.; Molina, Í. Detection of Crop Hail Damage with a Machine Learning Algorithm Using Time Series of Remote Sensing Data. *Agronomy* **2021**, *11*, 2078. <https://doi.org/10.3390/agronomy11102078>.
31. Busico, G.; Colombani, N.; Fronzi, D.; Pellegrini, M.; Tazioli, A.; Mastrocicco, M. Evaluating SWAT Model Performance, Considering Different Soils Data Input, to Quantify Actual and Future Runoff Susceptibility in a Highly Urbanized Basin. *J. Environ. Manag.* **2020**, *266*, 110625. <https://doi.org/10.1016/j.jenvman.2020.110625>.
32. Diana, S.R.; Purnama, S.M.; Dharma, G.; Sutrisnanto, A.; Perwitasari, I.; Farida, F. Estimation the Amount of Oil Palm Production Using Artificial Neural Network and NDVI SPOT-6 Imagery. *Int. J. Innov. Sci. Res. Technol.* **2019**, *4*, 548–554.
33. Hilal, Y.Y.; Ishak, W.; Yahya, A.; Asha'ari, Z.H. Development of Genetic Algorithm for Optimization of Yield Models in Oil Palm Production. *Chil. J. Agric. Res.* **2018**, *78*, 228–237. <https://doi.org/10.4067/S0718-58392018000200228>.
34. Kartika, N.D.; Astika, I.W.; Santosa, E. Oil Palm Yield Forecasting Based on Weather Variables Using Artificial Neural Network. *Indones. J. Electr. Eng. Comput. Sci.* **2016**, *3*, 626–633. <https://doi.org/10.11591/ijeecs.v3.i2.pp626-633>.
35. Moldovan, D.; Anghel, I.; Cioara, T.; Salomie, I. Time Series Features Extraction Versus LSTM for Manufacturing Processes Performance Prediction. In Proceedings of the 2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), Timisoara, Romania, 10–12 October 2019; IEEE: New York, NY, USA, 2019; pp. 1–10. <https://doi.org/10.1109/SPED.2019.8906653>.
36. Shareef, M.A.; Hasan, S.F. Characterization and Estimation of Dates Palm Trees in an Urban Area Using GIS-Based Least-Squares Model and Minimum Noise Fraction Images. *J. Ecol. Eng.* **2020**, *21*, 78–85. <https://doi.org/10.12911/22998993/123252>.
37. Das, A.C.; Noguchi, R.; Ahamed, T. An Assessment of Drought Stress in Tea Estates Using Optical and Thermal Remote Sensing. *Remote Sens.* **2021**, *13*, 2730. <https://doi.org/10.3390/rs13142730>.

5.2. Objetivo específico 3

El objetivo específico 3 consiste en:

OE 3: Determinar la relación entre los espectros de mejor ajuste y la edad fenológica del cultivo, condiciones climáticas y edáficas.

Sobre este objetivo se escribió el artículo:

Estimación de la edad del cultivo de palma de aceite en función de índices de vegetación y humedad generados a partir imágenes satelitales mediante técnicas de aprendizaje automático

El artículo fue enviado a la revista Alcances Tecnológicos (ISSN: 2215-5481), editada por el Instituto Nacional de Innovación y Transferencia en Tecnología Agropecuaria (INTA), ente adscrito al Ministerio de Agricultura y Ganadería (MAG) de Costa Rica. Es una publicación internacional de carácter académico y científico. Su formato y publicación es electrónico.

Referencia: Gómez-Calderón, N.; Watson-Hernández, F.; Quesada-Mora, Y; Rojas-Vasquez, R. (2022) Estimación de la edad del cultivo de palma de aceite en función de índices de vegetación y humedad generados a partir imágenes satelitales mediante técnicas de aprendizaje automático.

Estimación de la edad del cultivo de palma de aceite en función de índices de vegetación y humedad generados a partir imágenes satelitales mediante técnicas de aprendizaje automático

Estimation of oil palm crop age based on vegetation and humidity indices generated from satellite images using machine learning techniques

Natalia Gómez Calderón

Afiliación institucional: Instituto Tecnológico de Costa Rica, Escuela de Ingeniería Agrícola

Correo electrónico: ngomez@itcr.ac.cr

ORCID: <https://orcid.org/0000-0001-7961-7529>

Fernando Watson Hernández

Afiliación institucional: Instituto Tecnológico de Costa Rica, Escuela de Ingeniería Agrícola

Correo electrónico: fwatson@itcr.ac.cr

ORCID: <https://orcid.org/0000-0001-8258-4668>

Roger Rojas Vásquez

Afiliación institucional: Instituto Tecnológico de Costa Rica, Escuela de Ingeniería Agrícola

Correo electrónico: rogerrojas@estudiantec.cr

ORCID: <https://orcid.org/0000-0002-1607-3680>

Yerlin Quesada Mora

Afiliación institucional: Instituto Tecnológico de Costa Rica, Escuela de Ingeniería Agrícola

Correo electrónico: quesadayer@estudiantec.cr

ORCID: <https://orcid.org/0000-0001-6658-2494>

1. Resumen

El aceite de palma es uno de los aceites vegetales más consumidos en el mundo y es un ingrediente clave en cadenas de valor globales rentables. En Costa Rica, es uno de los tres cultivos con mayor área agrícola ocupada. Este estudio tiene como objetivo estimar de la edad del cultivo de palma aceitera en función de índices de vegetación y humedad generados con imágenes satelitales mediante técnicas de aprendizaje automático, para esto se consideró información de 28 años de un cultivo de palma aceitera ubicado en el Pacífico Central de Costa Rica y sus imágenes satelitales LANDSAT. Se generaron dos modelos, el modelo *Random Forest* (RF) presentó mejores resultados al usar un *n.tree* de 10000 en el entrenamiento, obteniéndose los siguiente coeficientes: NSE=0.9161, RMSE=1.6485 años, MAE=1.0178 años y $r^2=0.9476$; y el modelo *XGBoost* utilizando un *max.depth* de 4, *n.rouds* de 50 y *nthread* de 10 en el entrenamiento para el que se obtuvo: NSE=0.9530, RMSE=1.23 años, MAE=0.89 años y $r^2=0.9577$. Para los datos obtenidos en la validación, con RF resultaron los coeficientes de: NSE=0.5395, RMSE=3.76 años, MAE= 2.46 años y $r^2=0.5533$; para el modelo *XGBoost* se obtuvo NSE=0.6204, RMSE=3.4114 años, MAE= 2.2250 años y $r^2=0.6210$. La variable más importante en ambos modelos fue el percentil 75 del NDMI. El modelo más adecuado y con mejor rendimiento es *XGBoost*, sin embargo, no se recomienda cuando las edades del cultivo sobrepasen los 25 años.

Palabras Clave: *Random Forest*, *XGBoost*, *Landsat*, *Google Earth Engine*, agricultura digital

2. Abstract

Palm oil is one of the most widely consumed vegetable oils in the world and is a key ingredient in profitable global value chains. In Costa Rica, it is one of the three crops with the largest occupied agricultural area. The objective of this study is to estimate the age of the palm oil crop based on vegetation and humidity indices generated with satellite images using machine learning techniques, for this information will be taken from 28 years of a palm oil crop located in the Central Pacific of Costa Rica and its LANDSAT satellite images. Two models were generated, the Random Forest (RF) model presented better results when using a ntree of 10,000 in training, obtaining the following coefficients: NSE=0.9161, RMSE=1.6485 years, MAE=1.0178 years and $r^2=0.9476$; and the XGBoost model using a max.depth of 4, nrounds of 50 and nthread of 10 in the training for which it was obtained: NSE=0.9530, RMSE=1.23 years, MAE=0.89 years and $r^2=0.9577$. For the data obtained in the validation, the RF coefficients were: NSE=0.5395, RMSE=3.76 years, MAE= 2.46 years and $r^2=0.5533$; for the XGBoost model NSE=0.6204, RMSE=3.4114 years, MAE= 2.2250 years and $r^2=0.6210$. The most important variable in both models was the NDMI 75th percentile. The most suitable model with the best performance is XGBoost, however it is not recommended when the age of the crop exceeds 25 years.

Key Words: Random Forest, XGBoost, Landsat, Google Earth Engine, digital agriculture

3. Introducción

En Costa Rica uno de los principales cultivos agrícolas es la palma aceitera. Además, según Clare (2004) y Chowdhury et al. (2022), es una actividad rentable que mejora la economía de las zonas del mundo en las que se produce. Debido a su alto rendimiento en extracción de aceite, la palma se ha convertido en uno de los aceites vegetales más consumidos a nivel global (Chong et al. 2017), cuya plantación juega un papel vital en el mercado mundial de aceites comestibles (Chowdhury et al. 2022). Por lo anterior, es necesario encontrar una forma automática de predecir diversas variables de manejo del cultivo como una herramienta de apoyo para la toma de decisiones (Chowdhury et al. 2022).

Por otra parte, la palma de aceite es un árbol perenne, por lo que el cultivo tiene una estructura de dosel muy parecida a la de los bosques que otros cultivos agrícolas no presentan, por lo que se puede aplicar teledetección mediante imágenes aéreas o satelitales, esto según Morel (2012) y Rodríguez et al. (2019), que muestren tomas superiores de las áreas sembradas. Acorde con Rodríguez et al. (2019), con estas herramientas es posible estimar la densidad de árboles por subpíxeles de forma factible y con esto obtener mapas de densidad. El uso de imágenes satelitales para capturar toda la plantación es una herramienta adecuada para la detección de plantaciones y vegetación (Chowdhury et al. 2022).

Las imágenes multiespectrales brindan información que se obtiene de cada píxel, lo que permite estimar índices de vegetación (VIs) y realizar un análisis de mezcla espectral (SMA) (Morel, 2012). Los datos de cada píxel y del cultivo permiten determinar correlaciones con el uso de algoritmos para el análisis de datos, por lo que es recomendado evaluar más de un algoritmo de aprendizaje automático (*machine learning*) mejorando así la precisión de las estimaciones (Khanal et al. 2018), para lo que, según Amirrudin et al. (2020) y Zheng et al (2021) y se divide la base de datos en relación 7:3 o 8:2 para entrenamiento y validación respectivamente.

El manejo de datos mediante técnicas de aprendizaje automático ha demostrado buenos resultados y es muy prometedor para la evaluación del estado del cultivo de palma de aceite (Amirruddin et al. 2020), y a partir de los resultados se conoce que características son sensibles a las distorsiones introducidas por las imágenes (Chowdhury et al. 2022). Además, los datos basados en reflectancia espectral se han utilizado ampliamente en monitoreo de condiciones de los cultivos y estimación de su rendimiento (RYU et al. 2021). Sin embargo, en cuanto a estimación de edad del cultivo de palma aceitera no hay antecedentes.

En plantaciones forestales, la edad no se puede medir directamente haciendo uso de datos de detección remota, pero se puede inferir mediante variaciones en la reflectancia espectral que reflejan cambios en los factores forestales en el tiempo (Jakubauskas y Price, 2000). La predicción de la edad de plantaciones forestales depende de parámetros estructurales, tales como altura, área basal y densidad; y de los factores biofísicos como biomasa e índice de área foliar (LAI) que cambian con la edad (Lencinas y Mohr-Bell, 2007), que también pueden ser utilizados en el cultivo de palma aceitera.

Las imágenes LANDSAT 5 Thematic Mapper han sido utilizadas para obtener datos de radiancia y desarrollar un modelo de predicción de edad de un bosque de pino, mediante regresión múltiple escalonada, obteniendo una buena correlación entre la edad y los factores estructurales del bosque que controlan la respuesta espectral (Jakubauskas y Price, 2000). Asimismo, han sido utilizadas para estimar la edad de plantaciones de pino y eucalipto a través de ecuaciones generadas a partir de la correlación entre la edad y las variables explicativas (valores espectrales) con errores de 2.16 años para eucalipto y 3.52 años para pino (Lencinas y Mohr-Bell 2007).

Por otro lado, las imágenes satelitales multiespectrales LANDSAT son de acceso libre, suministran información de hasta 11 longitudes de onda (bandas), con resoluciones de entre 15 m y 100 m, rangos aceptables debido a que las zonas de manejo del cultivo de palma aceitera en Costa Rica superan las 4 hectáreas según datos del INEC (2015). Es importante encontrar relaciones entre la multiespectralidad de las imágenes, rendimiento, condiciones climáticas, edáficas y la edad del cultivo para la gestión de la producción de palma de aceite en función de índices de vegetación y humedad generados a partir imágenes LANDSAT 5, 7 y 8 en el Pacífico Central de Costa Rica, debido a la variabilidad climática de la región, con el fin de generar una herramienta que facilite la proyección del comportamiento del cultivo. El objetivo de este proyecto fue estimar de la edad del cultivo de palma de aceite en función de índices de vegetación y humedad generados a partir imágenes satelitales mediante técnicas de aprendizaje automático.

4. Metodología

En el estudio se consideraron ciento tres unidades productivas (103 UP) de una agroindustria de palma aceitera, ubicada en el Pacífico Central de Costa Rica. Se utilizó información histórica de 16-20 años del cultivo sobre variedad genética, rendimiento (t/ha) y año de siembra.

Para las unidades productivas consideradas, se obtuvieron las imágenes LANDSAT 5, 6 y 8 Collection 1 Tier 1 *calibrated top-of-atmosphere (TOA) reflectance* (Chander et al. 2009), desde enero de 1996 a diciembre del 2016. En la plataforma *Google Earth Engine* (GEE), fue creado un código de programación en lenguaje *Java Scrip*, para generar de forma cíclica (para cada imagen obtenida), 12 índices tanto de vegetación como de humedad, lo que generó una nueva colección de 20 años de historial con una resolución interanual para cada uno de los índices utilizados. La herramienta requirió como entrada la información de la geometría del área de

cada una de las 103 UP, las cuales constituyen la unidad mínima de área de información de producción *-feature collection-*, para obtener una descripción estadística del comportamiento de los índices en cada UP para cada fecha de captura de la imagen, determinando valores mínimos, máximos, cuartiles 25 y 75, medias, medianas y desviaciones estándar. Los valores estadísticos se ordenaron de forma cronológica y creando un archivo de forma automática para cada una de las UP. Cada UP contiene en total 84 variables.

Se obtuvieron nueve índices de vegetación y tres índices de agua, utilizando la distribución de las imágenes de satélite a través de canales de color RGB. El paso de cada banda por estos canales dota de diferentes tonalidades los elementos dependiendo de la respuesta de estos frente a las longitudes de onda (Ramos 2021). Los satélites LANDSAT 5, 7 y 8 distribuyen las señales recibidas en distintas bandas, su distribución denotada según sus siglas o nombre en inglés se muestra en los Cuadros 1 y 2.

Cuadro 1. Señales que reciben las distintas bandas para satélites LANDSAT 5, 7 y 8
Table 1. Signals received by the different bands for LANDSAT 5, 7 and 8 satellites

Landsat 5 y 7		Landsat 8	
Banda	Señal	Banda	Señal
1	BLUE	2	BLUE
2	GREEN	3	GREEN
3	RED	4	RED
4	NIR	5	NIR
5	SWIR 1	6	SWIR 1
7	SWIR 2	7	SWIR 2

Fuente: Earth Observing System (1999, 2013).

Cuadro 2. Índices de vegetación obtenidos a partir de datos multispectrales de las imágenes.

Table 2. Vegetation indices obtained from multispectral image data.

Variable	Índice	Ecuación*	Fuente
Vegetación	ARVI	$(NIR - (2 \cdot RED) + BLUE) / (NIR + (2 \cdot RED) + BLUE)$	[24]
Vegetación	AVI	$[NIR \cdot (1 - RED) \cdot (NIR - RED)]^{1/3}$	[25]
Vegetación	EVI	$2,5 \cdot (NIR - RED) / (NIR + 6 \cdot RED - 7,5 \cdot BLUE + 1)$	[26]
Vegetación	GCI	$(NIR / GREEN) - 1$	[24]
Vegetación	GNDVI	$(NIR - GREEN) / (NIR + GREEN)$	[26]
Vegetación	NDVI	$(NIR - RED) / (NIR + RED)$	[26]
Vegetación	NPCRI	$(RED - BLUE) / (RED + BLUE)$	[25]
Vegetación	SAVI	$1,5 \cdot (NIR - RED) / (NIR + RED + 0,5)$	[26]
Vegetación	SIPI	$(NIR - BLUE) / (NIR + BLUE)$	[26]
Agua	MSI	$(SWIR / NIR)$	[27]
Agua	NDMI	$(NIR - SWIR) / (NIR + SWIR)$	[27]
Agua	NDWI	$(GREEN - NIR) / (GREEN + NIR)$	[27]

En el software RStudio (Sykas 2020) se reordenó la base de datos que contenía edad del cultivo, material genético, lote, rendimiento e índices, eliminando los datos faltantes y generando una nueva base de datos a partir de esta. La cual contiene solo los índices y edad como la variable dependiente. Se filtraron los datos, eliminando los valores de $ndmi_mean < 0.2$ debido a que se estimaron como valores afectados por nubosidad que disminuyen el valor real del lote. En la Figura 1 se muestra un diagrama de la metodología utilizada.

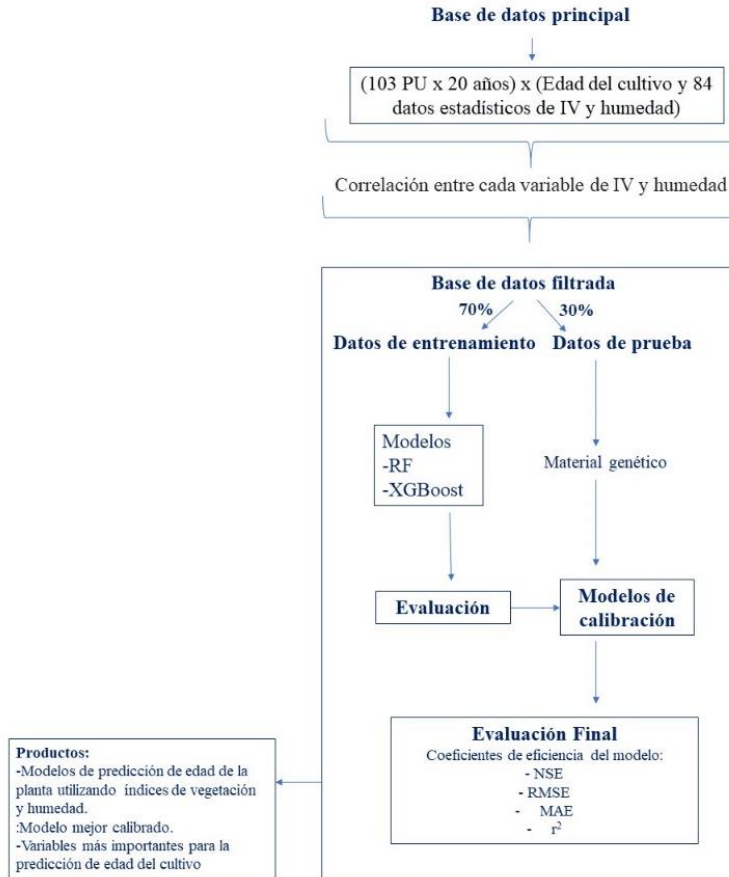


Figura 1. Diagrama metodológico del análisis de la información y sensibilidad de los modelos generados

Figure 1. Methodological diagram of the analysis of the information and model's sensitivity

La base de datos luego fue dividida en etapa de entrenamiento (Training Data) y validación (Test Data) de manera aleatoria, con 70 y 30% de los datos respectivamente. Se aplicaron los modelos de *Random Forest (RF)* y *XGBoost*, cuyos resultados fueron evaluados aplicando cuatro coeficientes para obtener la mejor calibración de los modelos, a continuación, su nombre y abreviatura según sus siglas en inglés: Coeficiente de eficiencia de Nash-Sutcliffe (NSE), la raíz del error cuadrático medio (RMSE), el coeficiente medio de error absoluto (MAE) y el coeficiente de determinación (r^2), mostrados en el Cuadro 3.

Cuadro 3. Coeficientes para determinación de la eficiencia de los modelos resultantes.
 Table 3. Coefficients to determine the efficiency of the resulting models.

Coeficiente	Ecuación*
NSE	$1 - \frac{\sum_{i=1}^n (y_i^{obs} - y_i^{sim})^2}{\sum_{i=1}^n (y_i^{obs} - \bar{y}^{obs})^2}$
RMSE	$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i^{obs} - y_i^{sim})^2}$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i^{obs} - y_i^{sim}|$$

$$r^2 = \left(\frac{\sum [(y_i^{obs} - \bar{y}^{obs})(y_i^{sim} - \bar{y}^{sim})]}{\sqrt{\sum (y_i^{obs} - \bar{y}^{obs})^2 * \sum (y_i^{sim} - \bar{y}^{sim})^2}} \right)^2$$

*Donde n es la cantidad de datos disponibles para el estudio, y_i^{obs} corresponde a la información de la edad del cultivo obtenido para el intervalo en estudio, \bar{y}^{obs} es el promedio de la edad del cultivo obtenido para el intervalo en estudio, y_i^{sim} es la edad del cultivo simulada y \bar{y}^{sim} es el promedio de la edad del cultivo simulada.

5. Resultados

La plantación de palma utilizada se caracteriza por tener una edad media de 10.35 años con un rango máximo de edad entre 0 y los 32 años, pero en donde, la gran mayoría de los datos se concentran entre los 3 y 19 años. La información de la edad de la planta de palma aceitera utilizada tiene un comportamiento normal (Figura 2).

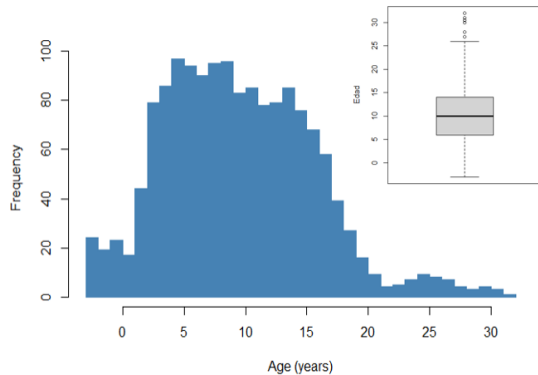


Figura 2. Caracterización grafica del comportamiento de la edad de la palma de aceita.
Figure 2. Characterization of oil palm´s age.

5.1. Coeficientes de los modelos de predicción de la edad del cultivo

Se construyeron modelos de aprendizaje automático (Machine Learning) según las estructuras de *RF* y *XGBoost*, los resultados obtenidos en la calibración de los modelos se muestran en el Cuadro 4. Los valores de NSE y r^2 se clasificaron según los rangos de aceptabilidad expuestos por Moriasi et al. (2007). Para RMSE y MAE se consideró que errores menores al 25% de la edad media es aceptable y más de un 40% insatisfactorio.

Para el modelo RF, se probaron 3 diferentes valores de número de árboles (*ntree*): 500, 5000 y 10000. El mejor desempeño según los coeficientes de evaluación estadística utilizados se obtuvo para *ntree* = 10000. En el modelo *XGBoost*, se utilizó un *max.depth* de 4, *nrounds* de 50 y *nthread* de 10. Se filtraron los datos por correlación entre índices y edad del cultivo. En ambos modelos se utilizó además de los VIs, una clasificación numérica del material genético.

En la etapa de entrenamiento ambos modelos (*RF*, *XGBoost*) obtienen valores de los coeficientes NSE y r^2 superiores a 0.91, con MAE promediado de 0.95 años y RMSE promedio de 1.44 años. Estos rendimientos disminuyen en la validación, en donde el modelo *XGBoost* tiene mejores valores en los índices de eficiencia y error que el modelo RF (cuadro 4).

Cuadro 4. Coeficientes de evaluación de calibración de los modelos utilizados.
 Table 4. Calibration evaluation coefficients of the models used.

Modelo	Fase	NSE	MAE	RMSE	r ²
RF	Train	0.9161	1.0178	1.6485	0.9476
	Test	0.5395	2.4589	3.7571	0.5533
XGBoost	Train	0.9530	0.8859	1.2333	0.9577
	Test	0.6204	2.2250	3.4114	0.6210

EL XGBoost en la validación obtuvo r²=0.62 y un NSE=0.62, valores que según Moriasi et al. (2007) son aceptables y los coeficientes de error MAE=2.22 años y RMSE=3.4 años, este último excede el valor establecido como aceptable (2.6 años).

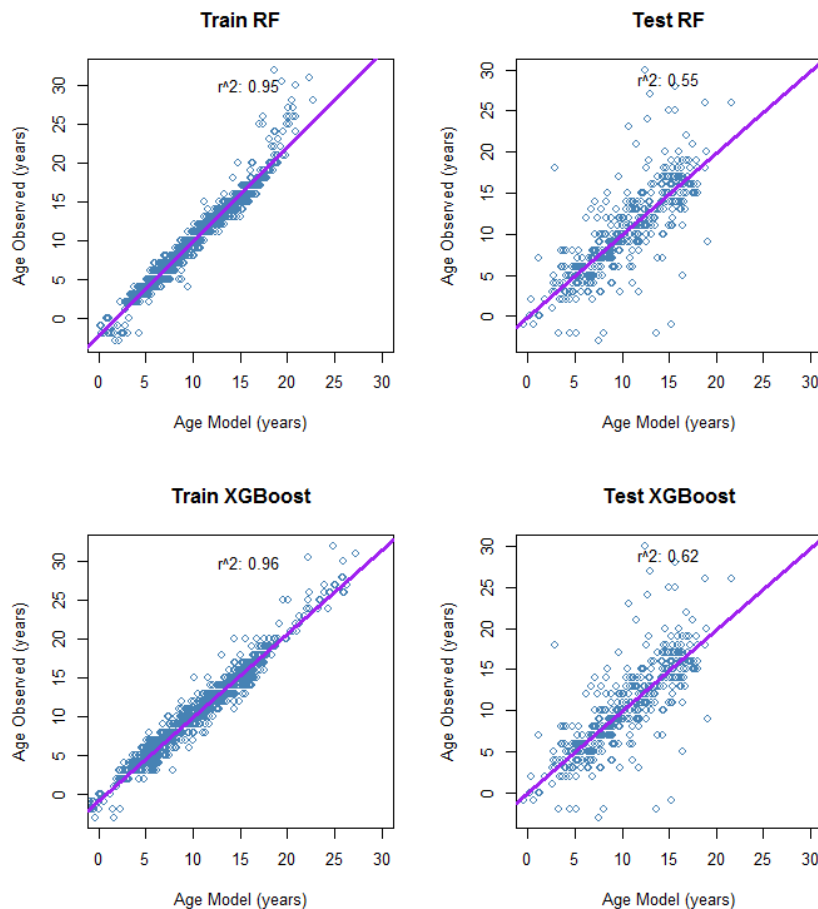


Figura 3. Comparación de la edad observada con la edad simulada, en los procesos de calibración y validación.

Figure 3. Comparison of the observed age with the simulated age, in the calibration and validation processes.

Sin embargo, los modelos muestran un mejor desempeño en el rango de edades en donde mayor información existe, lo cual es de esperar, debido a que cuentan con un número mayor de datos para su entrenamiento. Los principales problemas de modelos se observan al identificar edades que superen los 20 años (Chemura et al. 2015) tendiendo a subestimar la edad, tal como se muestra en la figura 4, en donde destaca un lote con edad superior a los 28 años y los modelos les asignan una edad inferior a los 4 años. En tanto, para edades entre los 8 y 18 años los modelos muestran un buen desempeño comparado con otros trabajos (Tan et

al. 2013). Al acotar el rango de edades entre 0 – 20 años ambos modelos obtienen en promedio NSE=0.81, RMSE=2.14 años, MAE=1.6 años y $r^2=0.80$.

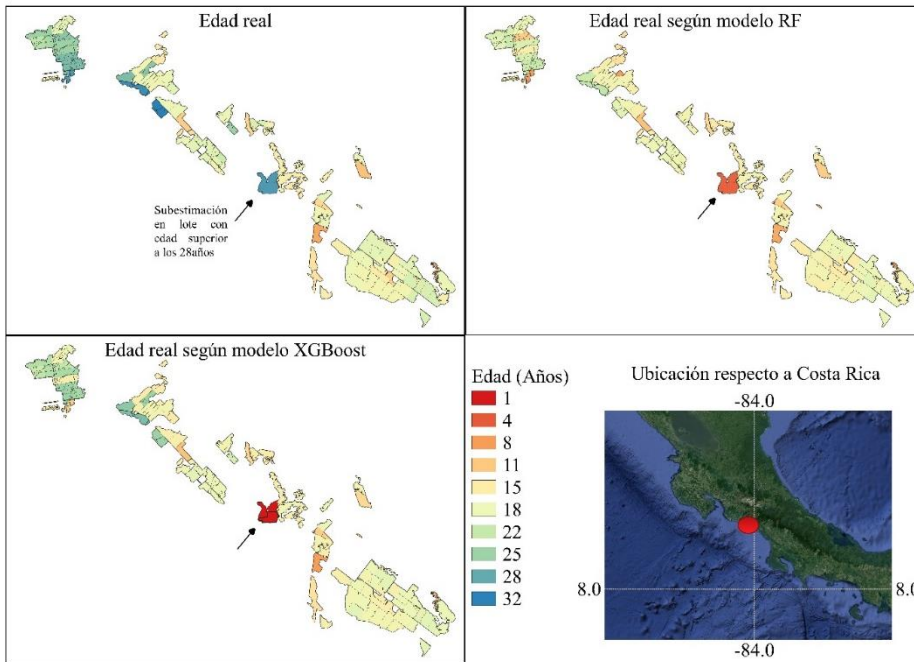


Figura 4. Visualización de resultados para el año 2016

5.1. Importancia de los índices de vegetación y humedad para la predicción de la edad del cultivo

El índice más significativo para ambos modelos fue el `ndmi_p75` (Figura 5). Para el modelo RF la variable MSI es la segunda en importancia y la variación espacial de estos dos VIs también es utilizada (McMorrow 2001), ya que variables de los estadísticos p25, p75, max y min aparecen ente los 10 mejor calificados. Tanto, en el modelo `XGBoost` la segunda variable en importancia el `npcri_min` su importancia relativa ronda el 25%, por lo que, en este modelo el principal peso estará en la variabilidad del percentil 75 del NDMI p75 (Figura 5).

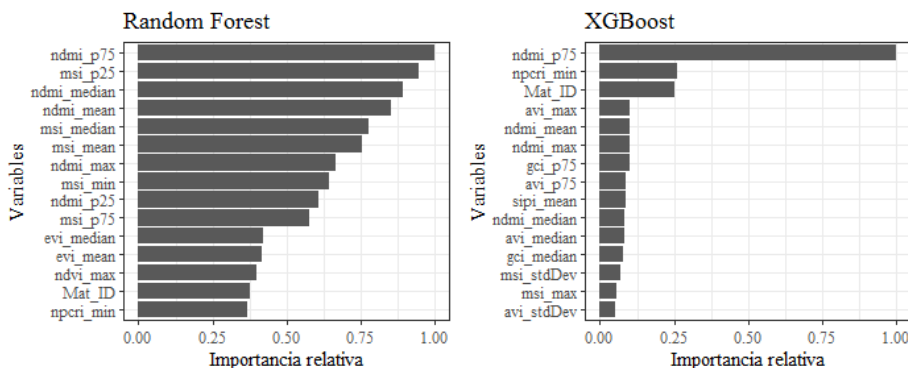


Figura 5. Importancia relativa de los índices en cada uno de los modelos

Los resultados obtenidos concuerdan con diversos autores, demostrando que el NDVI, SAVI y otros índices de vegetación tienen poca o nula correlación con la edad de la palma (Tan et

al. 2013). Además, se demuestra que para las condiciones de nubosidad del Pacífico de Centroamérica es posible determinar la edad de unidades productivas de palma de aceite por medio de la utilización de imágenes satelitales Landsat. Sin embargo, el estudio presenta limitaciones, ya que, tienen problemas para la identificación áreas de cultivo con edades superiores a los 20 años, lo que probablemente se debe a poca información o que después de esa edad la variación de las copas de la palma es poca y no permite realizar estimaciones por este método (Chemura et al. 2015).

La utilización de modelos de elevación para utilizar como variable la altura del cultivo, podría mejorar este tipo de modelos de predicción, entendiendo que otros estudios han generado mejores resultados a escalas menores de área y utilizando información LiDAR dentro de sus variables (Rizeei et al. 2018).

6. Conclusiones

En la evaluación y predicción de edad para la planta de palma de aceite con cualquier material genético analizado utilizando técnicas de aprendizaje automático *XGBoost* y *Random Forest*, además de variables de IV y humedad obtenidos como producto del análisis de imágenes satelitales Landsat 6, 7 y 8. El modelo más adecuado y con mejor rendimiento es *XGBoost*.

De los índices de vegetación y humedad de mayor importancia en ambos modelos los más destacados son los índices de humedad NDMI y MSI, para *Random Forest*, estos índices tienen mayor representación en sus percentiles 25 y 75 y aunque también están presentes entre las variables más importantes para *Random Forest*, en el *XGBoost*, que fue el mejor valorado, la variable más importante es el percentil 75 del índice NDMI, significando que para la predicción y evaluación de la edad para la planta de palma de aceite en la zona del Pacífico Central de Costa Rica se deben considerar la distribución de los datos dentro de las unidades productivas (UP) para optimizar el desempeño de los modelos.

7. Agradecimientos

A los productores de palma aceitera del Pacífico Central de Costa Rica por facilitar los registros históricos de rendimientos de la producción y a la Vicerrectoría de Investigación y Extensión del Tecnológico de Costa Rica por brindar las facilidades para desarrollar el estudio “Modelación del rendimiento de cultivo de palma de aceite a partir de imágenes satelitales multiespectrales, en el Pacífico Central de Costa Rica”, proyecto inscrito en el año 2020.

8. Referencias bibliográficas

- Amirruddin, AD; Muharam, FM; Ismail, MH; Tan, NP; Ismail, MF. 2020. Hyperspectral spectroscopy and imbalance data approaches for classification of oil palm's macronutrients observed from frond 9 and 17. *Computers and Electronics in Agriculture* 178:105768. DOI: 10.1016/j.compag.2020.105768.
- Chander, G; Markham, BL; Helder, DL. 2009. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors; *Remote Sensing of Environment*. 113(5):893–903. DOI: 10.1016/j.rse.2009.01.007.
- Chemura, A; van Duren, I; van Leeuwen, LM. 2015. Determination of the age of oil palm from crown projection area detected from WorldView-2 multispectral remote sensing data: The case of Ejisu-Juaben district, Ghana. *ISPRS Journal of Photogrammetry and Remote Sensing* feb. 2015. 100:118–127, DOI: 10.1016/j.isprsjprs.2014.07.013.

- Chong, KL; Kanniah, KD; Pohl, C; Tan, KP. 2017. A review of remote sensing applications for oil palm studies. *Geo-Spatial Information Science* 20(2):184–200. DOI: 10.1080/10095020.2017.1337317.
- Chowdhury, PN; Shivakumara, P; Nandanwar, L; Samiron, F; Pal, U; Lu, T. 2022. Oil palm tree counting in drone images. *Pattern Recognition Letters* 153:1-9. DOI: 10.1016/j.patrec.2021.11.016.
- Clare P. 2004. El cultivo de la palma aceitera en Costa Rica en el contexto del TLC con los Estados Unidos de Norteamérica. *Revista Espiga* 5(9):95-124. DOI: 10.22458/re.v5i9.1108.
- Dube, T; Pandit, S; Shoko, C; Ramoelo, A; Mazvimavi, D; Dalu, T. 2019. Numerical assessments of leaf area index in tropical savanna rangelands, South Africa using Landsat 8 OLI derived metrics and in-situ measurements. *Remote Sensing*. 11(7). DOI: 10.3390/rs11070829.
- Earth Observing System. 1999. LANDSAT 7.
- Earth Observing System. 2012. LANDSAT 5 (TM).
- Earth Observing System. 2013. LANDSAT 8.
- INEC. 2015. VI Censo Nacional Agropecuario: Características de las Fincas y de las Personas Productoras. San José, Costa Rica.
- Jakubauskas, ME; Price, K. 2000. Regression-Based Estimation of Lodgepole Pine Forest Age from Landsat Thematic Mapper Data. *Geocarto International* 15(1):21–26. DOI: 10.1080/10106040008542136.
- Khanal, S; Fulton, J; Klopfenstein, A; Douridas, N; Shearer, S. 2018. Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. *Computers and Electronics in Agriculture* 153:213–225. DOI: 10.1016/j.compag.2018.07.016.
- Lencinas, JD; Mohr-Bell, D. 2007. Estimación de clases de edad de las plantaciones de la provincia de Corrientes, Argentina, con base en datos satelitales Landsat. *Bosque, Valdivia*. 28(2). DOI: 10.4067/S0717-92002007000200003.
- McMorrow, J. 2001. Linear regression modelling for the estimation of oil palm age from Landsat TM. *International Journal of Remote Sensing*. 22(12):2243–2264. DOI: 10.1080/01431160117188.
- Morel, AC; Fisher, JB; Malhi, Y. 2012. Evaluating the potential to monitor aboveground biomass in forest and oil palm in Sabah, Malaysia, for 2000–2008 with Landsat ETM+ and ALOS-PALSAR. *International Journal of Remote Sensing* 33(11):3614–3639. DOI: 10.1080/01431161.2011.631949.
- Moriasi, D; Arnold J; Van-Liew, M; Bingner, R; Harmel, R; Veith, T. 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE*. 50(3):885–900. DOI: 10.13031/2013.23153.
- Ramos, D. 2021. Combinación de bandas en imágenes de satélite Landsat y Sentinel. *MappingGis*.
- Rizeei, HM; Shafri, HZM; Mohamoud, MA; Pradhan, B; Kalantar, B. 2018. Oil Palm Counting and Age Estimation from WorldView-3 Imagery and LiDAR Data Using an Integrated OBIA Height Model and Regression Analysis. *Journal of Sensors*. 2018:1–13. DOI: 10.1155/2018/2536327.

- Rodríguez, AC; D'Aronco, S; Schindler, K; Wegner, JD. 2021. Mapping oil palm density at country scale: An active learning approach. *Remote Sensing of Environment* 261:112479. DOI: 10.1016/j.rse.2021.112479.
- Rodríguez, A; Ruiz-Ramos, M; Palosuo, T; Carter, TR; Fronzek, S; Lorite, IJ; Ferrise, R; Pirttioja, N; Bindi, M; Baranowski, P; Buis, S; Cammarano, D; Chen, Y; Dumont, B; Ewert, F; Gaiser, T; Hlavinka, P; Hoffmann, H; Höhn, JG; Jurecka, F; Kersebaum, KC; Krzyszczak, J; Lana, M; Mechiche-Alami, A; Minet, J; Montesino, M; Nendel, C; Porter, JR; Ruget, F; Semenov, MA; Steinmetz, Z; Stratonovitch, P; Supit, I; Tao, F; Trnka, M; de Wit, A; Rötter, RP. 2019. Implications of crop model ensemble size and composition for estimates of adaptation effects and agreement of recommendations. *Agricultural and Forest Meteorology* 264:351–362. DOI: 10.1016/j.agrformet.2018.09.018.
- RStudio. 2018. R Studio. The R Foundation for Statistical Computing. Boston, MA.
- RYU, J-H; OH, D; CHO, J. 2021. Simple method for extracting the seasonal signals of photochemical reflectance index and normalized difference vegetation index measured using a spectral reflectance sensor. *Journal of Integrative Agriculture* 20(7):1969–1986. DOI: 10.1016/S2095-3119(20)63410-4.
- Suharjito; Elwirehardja, GN; Prayoga, JS. 2021. Oil palm fresh fruit bunch ripeness classification on mobile devices using deep learning approaches. *Computers and Electronics in Agriculture* 188:106359. DOI: 10.1016/j.compag.2021.106359.
- Sykas, D. 2020. Remote Sensing Satellite Data and Spectral Indices. *Spectral Indices with multispectral satellite data*.
- Tan, C; Zhou, X; Zhang, P; Wang, Z; Wang, D; Guo, W; Yun, F. 2020. Predicting grain protein content of field-grown winter wheat with satellite images and partial least square algorithm. *PLoS ONE*. 15(3):1–15. DOI: 10.1371/journal.pone.0228500.
- Tan, KP; Kanniah, KD; Cracknell, AP. 2013. Use of UK-DMC 2 and ALOS PALSAR for studying the age of oil palm trees in southern peninsular Malaysia. *International Journal of Remote Sensing* oct. 2013. 34(20):7424–7446. DOI: 10.1080/01431161.2013.822601.
- Yousefi D.B., M; Mohd Rafie, AS; Abd Aziz, S; Azrad, A; Mazmira Mohd Masri, M; Shahi, A; Marzuki, OF. 2020. Classification of oil palm female inflorescences anthesis stages using machine learning approaches. *Information Processing in Agriculture*. DOI: 10.1016/j.inpa.2020.11.007.
- Zakir, M. 2018. GIS and RS Based Spatio-Temporal Analysis of Soil Moisture/Water Content Variation in Southern Irrigated Part of Sindh, Pakistan. *Journal of Remote Sensing & GIS*. 07(04). DOI: 10.4172/2469-4134.1000250.
- Zheng, J; Fu, H; Li, W; Wu, W; Yu, L; Yuan, S; William Tao, WY; Pang, TK; Kanniah, KD. 2021. Growing status observation for oil palm trees using Unmanned Aerial Vehicle (UAV) images. *ISPRS Journal of Photogrammetry and Remote Sensing* 173:95–121. DOI: 10.1016/j.isprsjprs.2021.01.008.

5.3. Objetivo específico 4

El objetivo específico 4 consiste en:

OE 4: Evaluar la relación entre el rendimiento de la palma aceitera con variables físicas y biológicas, mediante el uso del Modelo de Simulación de Sistemas de Producción Agrícola (APSIM).

Sobre este objetivo se escribió la tesis de maestría del Ing. Fernando Watson Hernandez:

Cuantificación y valuación de la aplicación de riego en el cultivo de palma de aceite como medida adaptativa al cambio climático por medio de la utilización de modelos de simulación de producción agrícola (apsim) calibrados para condiciones del Pacífico Central de Costa Rica

La tesis corresponde al Máster en cambio climático de la Universidad Internacional Iberoamericana. La tesis está en proceso de revisión por parte de la universidad. Se adjunta el documento completo, en el formato solicitado para su presentación.

MEDIO AMBIENTE

Cuantificación Y Valuación De La Aplicación De Riego En El Cultivo De Palma De Aceite Como Medida Adaptativa Al Cambio Climático Por Medio De La Utilización De Modelos De Simulación De Producción Agrícola (APSIM) Calibrados Para Condiciones Del Pacífico Central De Costa Rica

Máster En Cambio Climático

Fernando Alberto Watson Hernández

CRMAMCC1960925

CARTAGO, COSTA RICA

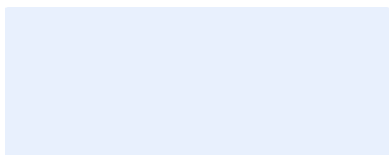
2022

COMPROMISO DE AUTOR

Yo, Fernando Watson Hernández, declaro que:

El contenido del presente documento es original y constituye un reflejo de mi trabajo personal. Manifiesto que, ante cualquier notificación de plagio, autoplagio, copia o falta a la fuente original, soy responsable directo legal, económico y administrativo sin afectar al Director del trabajo, a la Universidad y a cuantas instituciones hayan colaborado en dicho trabajo, asumiendo las consecuencias derivadas de tales prácticas.

Firma:



AUTORIZACIÓN PUBLICACIÓN ELECTRÓNICA

Cartago, 2022

Att: Dirección Académica

Por este medio autorizo la publicación electrónica de la versión aprobada de mi Proyecto Final bajo el título “Cuantificación y valuación de la aplicación de riego en el cultivo de palma de aceite como medida adaptativa al cambio climático por medio de la utilización de modelos de simulación de producción agrícola (apsim) calibrados para condiciones del pacífico central de Costa Rica” en el campus virtual y en otros espacios de divulgación electrónica de esta Institución.

Informo los datos para la descripción del trabajo:

Título	Cuantificación y valuación de la aplicación de riego en el cultivo de palma de aceite como medida adaptativa al cambio climático por medio de la utilización de modelos de simulación de producción agrícola (APSIM) calibrados para condiciones del pacifico central de Costa Rica
Autor	Fernando Watson Hernández
Resumen	
Programa	Máster En Cambio Climático
Palabras clave	Cambio Climático, Palma de Aceite, Riego, APSIM
Contacto	fwatson@itcr.ac.cr

Atentamente,

Firma:

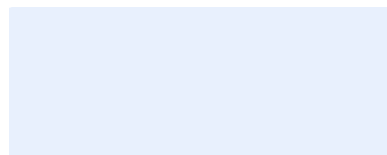


Tabla de contenido 2

1. Código y Título del Proyecto	3
2. Autores y direcciones	3
3. Resumen general del proyecto.....	3
4. Palabras clave.....	3
5. Resultados por objetivo	4
5.1. Objetivos específicos 1 y 2.....	4
5.2. Objetivo específico 3.....	22
1. Resumen.....	23
2. Abstract.....	24
3. Introducción.....	24
4. Metodología.....	25
5. Resultados.....	28
6. Conclusiones.....	31
7. Agradecimientos.....	31
8. Referencias bibliográficas	31
5.3. Objetivo específico 4.....	34
1. INTRODUCCIÓN.....	4
2. MARCO TEÓRICO.....	5
3. METODOLOGÍA	28
4. RESULTADOS.....	33

5. DISCUSIÓN	51
6. CONCLUSIONES Y RECOMENDACIONES.....	53
7. BIBLIOGRAFÍA.....	54
8. ANEXOS.....	61

RESUMEN

El cambio climático presenta escenarios contraproducentes para la producción agrícola. Generando especial interés en el estudio de medidas de adaptación que eviten establecer prácticas erróneas, debido a que sus efectos podrán ser más contraproducentes que los beneficios generados (Naciones Unidas 2017).

El presente estudio evalúa el efecto de la aplicación de riego al cultivo de palma de aceite como medida adaptativa al cambio climático por medio de la utilización de modelos de simulación de producción agrícola calibrados para condiciones del Pacífico Central de Costa Rica.

Para poder realizar lo planteado, se realizó un análisis clúster sobre el conjunto de unidad de producción, permitiendo identificar un grupo de lotes de trabajo representativos. Sobre este grupo se generaron simulaciones de sistemas de producción calibradas en el software APSIM. En estos modelos calibrados se corrieron tres escenarios; una Línea base para el periodo 2000-2019, un escenario de cambio climático para mediados de siglo del 2040 al 2059 (ECC1) y otro para finales de siglo del 2080 al 2099 (ECC2). La comparación de las corridas de Línea base contra ECC1 y ECC2 permitió cuantificar los efectos del cambio climático sobre el rendimiento del cultivo de palma de aceite. Finalmente, en las simulaciones los ECC 1 y 2 se aplicó irrigación como medida adaptativa. Determinando los valores de agua disponible en el suelo que mantenga los rendimientos generados en la Línea base. Se establece que para el ECC1 se tendrá una disminución de las lluvias acumuladas anuales del 5.55% y un aumento en la temperatura promedio de 1.73°C, variaciones que generan una disminución del rendimiento de la producción del cultivo de un 7.86%. Para el ECC2 las variaciones son mayores con una disminución de la precipitación del 18.06% y un aumento de la temperatura de 3.31°C, afectando el rendimiento en un 37.86% respecto de la producción media de la Línea Base (21.63 t/ha/año).

Se concluye que, según las simulaciones realizadas, el riego permite adaptar las condiciones de agua disponible en el suelo, al punto que mantener los rendimientos de la Línea base del cultivo de palma para los escenarios de cambio climático propuestos (ECC1 y ECC2). En ambos casos, en el análisis interanual mostro un incremento en los requerimientos hídricos del cultivo de palma de aceite conforme se establece la época seca. Además, que conforme el ECC se vuelva más intenso el agua disponible en el suelo disminuye aumentando los requerimientos de riego del cultivo de palma de aceite.

Palabras clave:

Cambio climático; Medidas de adaptación; Irrigación; APSIM; Palma de aceite

1. INTRODUCCIÓN

El cambio climático (CC) generará que los rendimientos de la producción de palma de aceite disminuyan paulatinamente en todo el mundo (Paterson, Kumar, Taylor, & Lima, 2015). Además, el CC podría movilizar hacia las regiones subtropicales las zonas más idóneas para el cultivo de Palma, debido a que los trópicos se volverán muy calientes (Paterson & Lima, 2018). Se estima que los efectos del CC en regiones tropicales serán aumento en la erosión de suelo, aumento en las temperaturas, mayor frecuencia de eventos extraordinarios de precipitación, pero con acumulados anuales menores y un aumento en la ocurrencia de sequías (Ghini, Bettioli, & Hamada, 2011). Según Suresh (2013) estos efectos del CC se traducen a un estrés hídrico en la palma de aceite, ocasionando una supresión la formación de inflorescencias femeninas aumentando el aborto, lo que afecta la relación entre sexos y el posterior desarrollo de las inflorescencias, lo que finalmente disminuye el rendimiento promedio de la plantación.

Algunos estudios realizados en palma de aceite han demostrado que, bajo condiciones de estrés hídrico, el riego permite aumentar los rendimientos del cultivo. Foo (1998) demostró que con la aplicación de riego se genera que al tercer año se aumente el rendimiento en 14 t/ha en comparación a zonas en donde no se aplica riego. Además, demostró que la curva de disminución natural del rendimiento del cultivo es menos intensa cuando existe irrigación que cuando no se aplica. Aun cuando estas investigaciones no se realizaron bajo un contexto de CC, permiten crear al menos la duda, si la aplicación de riego podría ser una medida para mantener los rendimientos ante un aumento de temperatura o disminución de la precipitación en el cultivo de palma de aceite.

En cultivos como la alcachofa se han realizado estudios demostrando que riego aplicando técnicas innovadoras contribuye a mitigar los problemas negativos relacionados con el cambio climático (Deligios et al., 2019). En regiones vinícolas también hay estudios que demuestran el potencial del riego de precisión como medida adaptativa al CC, sin embargo, los autores señalan, que estas acciones podrían generar una mayor demanda sobre un recurso que actualmente ya es escaso y por el cual se suscitan muchos conflictos sociales y políticos (Fraga, García de Cortázar Aauri, & Santos, 2018).

Por ejemplo, la cuenca del río Tempisque en Costa Rica, en donde el recurso hídrico es utilizado mayoritariamente para riego de cultivos agrícolas y por el cual existe una gran presión sobre todo en época seca principalmente en los meses de marzo y abril (Guzmán Arias & Calvo-Alvarado, 2012). Esta dinámica de uso intensivo en donde se aplican técnicas de riego con bajas eficiencias, aunado a problemas de desigualdad en el acceso y de retribución económica hacia los pobladores de la zona ha causado un clima de conflictividad (Alonso Ramírez Cover, 2007). Esta problemática podría aumentar producto del cambio climático y por la implementación de estrategias de adaptación sin una base técnica.

Las Naciones Unidas (2017) han advertido sobre la implementación de prácticas adaptativas erróneas, debido a que sus efectos colaterales son más contraproducentes que los beneficios que generen. Pero esta organización insiste en que la única forma de identificar procesos de adaptación eficientes es por medio de la investigación.

Es por tales razones que la presente propuesta, pretende evaluar la aplicación de láminas de riego en el cultivo de palma de aceite como medida adaptativa al cambio climático, por medio de la utilización de modelos de simulación de producción agrícola (APSIM) calibrados para condiciones del Pacífico Central de Costa Rica.

Objetivo general

Evaluar el efecto de la aplicación de riego en el cultivo de palma de aceite como medida adaptativa al cambio climático por medio de la utilización de modelos de simulación de producción agrícola calibrados para condiciones del Pacífico Central de Costa Rica.

Objetivos específicos

- Generar los modelos de simulación de producción agrícola calibrados para el periodo observado de tiempo (1996 al 2016) para un número de unidades productivas que representen el comportamiento del rendimiento del área de siembra en estudio.
- Determinar los impactos sobre el rendimiento del cultivo debido a los escenarios de cambio climático planteados para los periodos del 2040 al 2059 (ECC1) y del 2080 al 2099 (ECC2), teniendo como línea base el periodo del 2000 al 2019.
- Determinar el cambio necesario en el agua disponible en el suelo debido a la irrigación que permita mantener los rendimientos obtenidos en la Línea base.

2. MARCO TEÓRICO

2.1. Palma de aceite

La palma de aceite es un árbol tropical perenne; debido a su alto rendimiento y sus costes relativamente bajos de producción, se ha convertido en uno de los productos más populares, sobrepasando a otros aceites vegetales, y en uno de los usos de tierra más rentables en zonas tropicales húmedas (Meijaard et al., 2018; Barcelos et al., 2015).

Su nombre científico es *Elaeis guineensis* Jacq, y es una especie endémica del occidente y centro de África, siendo la más conocida de sus descendientes la *Deli dura* por ser la mayor fuente parental femenina en la producción de semillas, asimismo se han desarrollado programas de mejoramiento utilizando distintas estrategias y fuentes de polen *pisifera* como AVROS, Ekona, La Me, Yangambi, Ghana y Nigeria , lo cual innovó el mejoramiento y la producción de este cultivo

(Alvarado & Escobar, 2017; Ayala D. et al., 2017). Durante los últimos años se han impulsado los cultivos híbridos entre *E. guineensis* y la especie nativa americana llamada *Elaeis oleifera* (Kunth, Cortés) localizada en América Central y en regiones del norte de Sudamérica, logrando un crecimiento intermedio de la planta con una cosecha de bajo costo, así como una mejora en la resistencia genética sobre la enfermedad de Pudrición de cogollo (Billote, 2019).

En Costa Rica, la ASD (Agricultural Services and Development) es una empresa encargada del desarrollo de variedades de semillas y clonación de palma de aceite, inicialmente comenzó a suministrar variedades tradicionales como Deli x AVROS, Deli x Ekona y Deli x Yangambi, aunque nuevas variedades como Deli x Nigeria, Evolution, Deli x La Me, Tanzania x Ekona, Bemenda x Ekona, entre otros se utilizan actualmente; siendo el material de Ekona una de las más valoradas debido al tamaño y alto contenido de aceite de sus racimos (Rethinam & Murugesan, 2018; Escobar et al., 2000)

Las plantaciones comerciales de palma de aceite se cultivan con una densidad de 120 a 150 plantas por hectárea, sin que se sobrepongan las copas de las palmas y así maximizar la luz y la fotosíntesis. El ciclo de vida de la palma inicia con las semillas pregerminadas criadas en vivero durante 6 a 12 meses, donde posteriormente la plántula es sembrada en el campo a la densidad definitiva, entre los 2 y 3 años la planta se encuentra en su fase de inmadurez o de construcción de rendimiento, y es hasta alrededor de los tres años después de la plantación que la palma empieza a ser fructífera, alcanzado su fase de máxima producción unos 10 años después de la plantación, cuando el área de la hoja y el rendimiento aumentan lineal y constantemente, teniendo una vida media de 25 años (21 a 23 años productivos) después de lo cual la palma es muy alta, dificultando la cosecha, además de que el rendimiento decrece por la disminución de palmeras y por plagas y enfermedades (Woittiez, Wijk, Slingerland, Noordwijk, & Giller, 2017). La cosecha de los racimos frescos se realiza aproximadamente cada 10 a 14 días en una plantación madura (Meijaard et al., 2018).

Las raíces de la palma crecen tanto horizontal (hasta 6 m) como verticalmente (entre 1,5 y 5 m), siendo las raíces primarias las que se encargan de sostener estructuralmente la planta y las secundarias actúan como anclaje; las raíces se concentran en la capa superficial del suelo hasta 0,3 m de profundidad para captar el contenido de nutrientes (Intara, Nusantara, Supanjani, Caniogo, & Ekawita, 2018).

Al ser una especie vegetal tropical la palma de aceite requiere de condiciones cálidas y de alta humedad para aprovechar al máximo su capacidad fotosintética, por lo cual sus condiciones óptimas de temperatura fluctúan entre 24 y 28 °C y una radiación (Rad) solar diaria de 16 a 17 MJ/m², siendo su temperatura máxima media de 30 a 32 °C y mínima media de 21 a 24 °C, donde el mes más frío del año no puede ser menor a los 15 °C; además, necesita una alta pluviosidad, con una precipitación anual entre 2000 y 2500 mm, con un mínimo de 100 mm mensuales, no obstante, puede soportar lluvias de 4000 mm hasta un límite de 5000 mm al año en suelos con un buen drenaje (aunque aumenta la probabilidad de enfermedades), por el contrario, se han reportado casos de plantaciones con precipitación menores a 1000 mm anuales y hasta cinco meses de periodo seco (Pirker et al., 2016; Meijaard et al., 2018).

En lo que respecta a las condiciones del suelo es preferible que sean con poca grava y texturas con un buen drenaje, idealmente texturas como arcillas bien estructuradas, arcillo arenoso, franco arcilloso y franco arcilloso limoso, y con un buen nivel de materia orgánica que proporcione nitrógeno y que sean profundos, de esta manera el sistema radicular de la planta puede desarrollarse a un volumen de enraizamiento adecuado. Aunque la palma de aceite no es un cultivo muy exigente, este se puede ver afectado por suelos sulfatados, ácidos (pH < 4.0), mal drenados, salinos (CE > 1 S/m), muy erosionados y lixiviados, arenosos profundos, de cenizas volcánicas, lateríticos y deficientes de nutrientes (Corley & Tinker, 2016). En América los cultivos de palma de aceite predominan en suelos oxisoles y ultisoles (Woittiez et al., 2017).

En cuanto a las condiciones topográficas, las pendientes pronunciadas afectan el rendimiento de la palma, por cuanto hay un aumento de los costos de producción, mantenimiento y cosecha, adicionalmente se presentan una alta escorrentía superficial de los fertilizantes así como erosión de la capa superior del suelo, por lo que las condiciones óptimas son zonas planas con una inclinación de 0 a 4°, hasta un máximo de 16° (Pirker et al., 2016).

Para mantener una plantación de alto rendimiento de palma de aceite durante periodos prolongados se requiere de la aplicación de fertilizantes. Este cultivo requiere principalmente de Nitrógeno (N), Potasio (K), Fósforo (P), Magnesio (Mg) y Boro (B) (Tabla 2.1). De acuerdo con Tiemann et al. (2018) los factores más importantes para poder determinar el momento adecuado para la aplicación de fertilizantes son la movilidad de los nutrientes y los procesos del suelo asociados con el agua, ya que en el caso de los fertilizantes solubles de N y K, estos se pierden por eventos extremos de lluvia, lo cual puede evitarse al aplicar la fertilización en meses de precipitaciones moderadas o bajas.

Tabla 2.1. Colocación óptima de fertilizantes en las plantaciones de palma aceitera

Nutrientes	Fuente del fertilizante	Zona de preferencia	Razón
N	Urea Nitrato de amonio Sulfato de amonio	Encima del borde de los círculos de las palmas o de las pilas de frondas contiguas	Disminución de la competencia de la vegetación del sotobosque Reducción de la volatilización del N
K	Cloruro de potasio	Dentro de los círculos de las palmas (0 – 6 años) Alrededor de los círculos de las palmas (7 – 10 años) Fuera de los círculos de las palmas (> 10 años)	Incremento de la eficiencia en la absorción de nutrientes de parte de las raíces
P	Fosfato de roca Superfosfato triple Superfosfato simple	Sobre el espacio entre filas y el borde exterior de los círculos de las palmeras Sobre la pila de frondas	Mitigación de los efectos de acidificación de los fertilizantes de N Disminución de la erosión y la escorrentía por la vegetación del sotobosque
Mg	Kieserita Dolomita	Dentro de los círculos de las palmas (0 – 5 años) Alrededor y fuera de los círculos de las palmas (> 5 años)	Incremento de la eficiencia en la absorción de nutrientes de parte de las raíces
B	Borato	Dentro de los círculos de las palmas	Incremento de la eficiencia en la absorción de nutrientes de parte de las raíces

Nota. "Chapter four - Feeding the palm: A review of oil palm nutrition" por T.T. Tiemann, C.R. Donough, Y.L. Li, R. Härdter, R. Norton, H.H. Tao, R. Jaramillo, T. Satyanarayana, S. Zingore, y T. Oberthür, 2018, *Advances in Agronomy*, 152, p.220 (<https://doi.org/10.1016/bs.agron.2018.07.001>). Derechos reservados.

Los síntomas visuales de deficiencias nutricionales se observan en los sistemas foliares de la planta (Woittiez et al., 2017).

En zonas que presentan periodos de sequía o con bajas precipitaciones se requiere de la aplicación de riego para mantener el rendimiento del cultivo. De acuerdo con Carr (2011) por cada 100 mm que aumenta el déficit hídrico, el rendimiento se pierde aproximadamente un 10%, este déficit se produce cuando la evapotranspiración real es mayor a las precipitaciones.

El riego en palma de aceite se puede llevar a cabo superficialmente, mediante la inundación o el bloqueo del drenaje, siendo un método más sencillo y económico, no obstante, solo es viable en zonas planas o de baja pendiente, mientras que otros métodos suponen una inversión de bombas y tuberías, siendo el sistema más eficiente el de riego por aspersión (Corley & Tinker, 2016).

A partir del estudio realizado por Palat et al (2000, como se citó en Corley & Tinker, 2016) determinaron que la aplicación del riego no es completamente lineal (anteriormente se asumía una constante de 5 mm/día), obteniendo 3,9 ton/ha al aplicar 150 L de agua a la palma al día, mientras que si se aplicaban 300 L el rendimiento aumentaba solamente un 1,9 ton/ha más.

Actualmente, los niveles críticos de déficit hídrico en las distintas fases de desarrollo de la palma y los volúmenes óptimos de agua que deben aplicarse a ella permanecen sin definir (Woittiez et al., 2017), siendo el grado de estrés por déficit hídrico el punto decisivo para saber cuándo regar, el cual puede determinarse a partir de la dinámica hídrica en la curva de retención de agua en el suelo del cultivo y puntos de cantidad de agua (Corley & Tinker, 2016).

En lo que concierne al cambio climático y el cultivo de palma de aceite, Paterson & Lima (2018) indican que las condiciones climáticas óptimas para el cultivo de palma

se verán afectadas gradualmente para el 2030 y más aún para el 2100, perjudicando la producción de este cultivo; el cambio climático reducirá las áreas cultivadas actuales, extendiéndose a nuevas áreas (desplazamiento el cultivo hacia los polos), además las condiciones del cultivo por el estrés abiótico (como son la precipitación, temperatura, salinidad del suelo, dióxido de carbono, etc) y biótico (como las enfermedades, plagas, polinizadores, etc) se verán afectados negativamente.

Cabe agregar que en Malasia (uno de los mayores productores de palma de aceite junto con Indonesia) pronostica una disminución de un 30% en el rendimiento de la palma en el caso de que la temperatura aumente 2 °C por encima del nivel óptimo y las precipitaciones disminuyan un 10%, lo que evidencia la incertidumbre a la que se enfrentan los demás países productores de palma de aceite y el impacto que esto tendrá económicamente (Paterson & Lima, 2018).

2.2. Propiedades del suelo

El suelo es un sistema poroso compuesto por minerales, materia orgánica, agua y aire (Rai, Singh, & Upadhyay, 2017), algunas de las propiedades físicas y de capacidad de retención de agua del suelo se muestran en la Tabla 2.2.

Tabla 2.2 Propiedades físicas y de capacidad de retención de agua del suelo

Propiedad	Descripción	Parámetro en APSIM - Soil
Textura	<ul style="list-style-type: none"> Corresponde a la distribución granulométrica del suelo, según la proporción relativa de arena (2 – 0,02 mm), limo (0,02 – 0,002 mm) y arcilla (< 0,002mm). 	-
Densidad aparente	<ul style="list-style-type: none"> Es la masa del suelo seco por unidad de volumen. Permite estimar la capacidad total de almacenamiento de agua, la textura, la tasa de infiltración, la compacidad o el estado de aireación del suelo. La densidad aparente en arenas es baja (1,0 – 1,1 g/cm³), en limos es media (1,01 – 1,3 g/cm³) y en arcillas es alta (1,3 – 1,6 g/cm³). 	-
Capacidad de campo	<ul style="list-style-type: none"> Es el contenido de agua que permanece en el suelo después de que el exceso de agua se ha drenado por gravedad, retenida a 0,33 bar de presión de succión. En suelos permeables con textura fina y estructura uniforme el drenaje ocurre en 2 o 3 	<i>Drained upper limit (DULL)</i>

	días, mientras que en texturas gruesas se dura unas cuantas horas.	
	<ul style="list-style-type: none"> En este punto los contenidos de agua y aire en el suelo son los ideales para el crecimiento de los cultivos. 	
Punto de marchitez permanente	<ul style="list-style-type: none"> Es el contenido de humedad donde la planta no puede extraer el agua que se encuentra en el suelo, siendo retenida a una presión de 15 bar. provocando que esta muera En suelos de textura gruesa ($\approx 10\% - 15\% v/v$) se retienen menores cantidades de agua que en texturas finas ($\approx 26\% - 32\% v/v$). 	15Bar lower limit (LL15)
Capacidad de saturación	<ul style="list-style-type: none"> Se presenta cuando todos los poros del suelo están ocupados con agua, es decir, no queda aire en el suelo. 	Soil water saturation (SAT)
Suelo Seco	<ul style="list-style-type: none"> Es el contenido de agua más bajo en las capas superficiales como resultado de la evaporación. 	Soil air dry (AirDry)

Nota. Adaptado de (Rai et al., 2017) y " S-map parameters for APSIM " por I. Vogeler, L. Lilburne, T. Webb, R. Cichota, J. Sharp, S. Carrick, H. Brown, y V. Snow, 2022, *MethodsX*, 101632, (<https://doi.org/10.1016/j.mex.2022.101632>). Derechos reservados.

La densidad aparente es un factor clave a la hora de evaluar la calidad del suelo, puesto que este refleja de forma indirecta la porosidad del mismo (estos son los espacios vacíos del suelo), por lo que aquellos suelos que presentan una densidad aparente alta apuntan a una condición pobre del suelo que afecta el crecimiento de las raíces y la disminución de la aireación y la infiltración, asimismo, la densidad aparente está asociada a la textura del suelo, siendo los suelos finos los que poseen pocos y pequeños poros (Figura 2.1) (ORGANIZACIÓN DE LAS NACIONES UNIDAS PARA LA AGRICULTURA Y LA ALIMENTACIÓN, 2009).

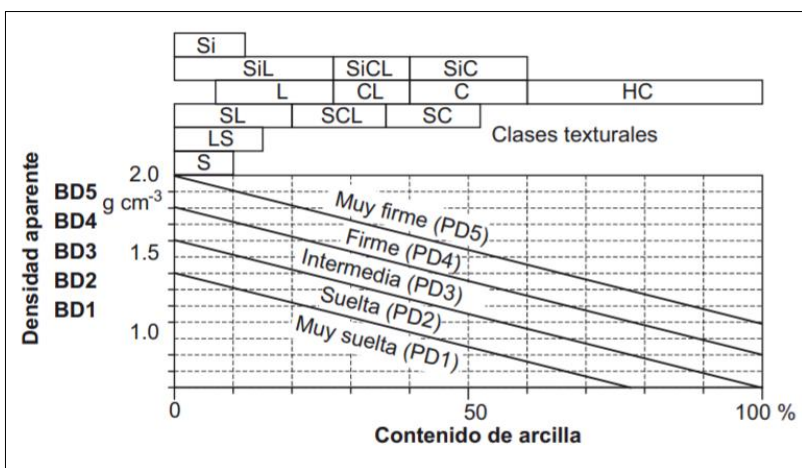


Figura 2.1 Calificación de la densidad aparente

Nota. Basado de Ad-hoc-AG-Boden (2005) como se citó en *Guía para la descripción de suelos* [Imágen] por Organización de las Naciones Unidas para la Alimentación y la Agricultura [FAO], 2009, FAO (<https://www.fao.org/3/a0541s/a0541s.pdf>). Derechos reservados.

El contenido de agua gravitacional que puede estar disponible para las plantas varía de acuerdo con el tipo del suelo, siendo los suelos arcillosos aquellos con mayor contenido de humedad que los suelos de texturas arenosas (Tabla 2.3), dado que las texturas más finas poseen una mejor agregación de sus partículas (mayor superficie y poros más pequeños y numerosos) y sus fuerzas de adhesión y cohesión son más fuertes, lo que también ocasiona que el agua este retenida en los microporos y las plantas no puedan acceder a ella; en cambio, las texturas gruesas poseen poros más grandes y una fuerza de adhesión y cohesión más baja provocando que el agua gravitacional fluya más rápido a través de los macroporos; es por ello que los suelos franco y franco limosos al tener un gran contenido de humedad disponible para las plantas son los más valorados en el área agrícola (Easton & Bock, 2016).

Tabla 2.3 Contenido promedio de agua disponible según la textura del suelo

Textura	Punto de marchitez permanente	Capacidad de campo	Contenido de agua disponible
	Porcentaje de humedad		
Arena	5	12	7
Franco arenoso	9	21	12
Franco	16	36	20
Franco limoso	18	39	21
Franco arcilloso	24	39	15
Limo arcilloso	24	39	13
Arcilloso	27	39	12

Nota. “ Soil and soil water relationships” por Z.M. Easton y E. Bock, 2016, Virginia Cooperative Extension, p.4 (<https://vtechworks.lib.vt.edu/handle/10919/75545>). Derechos reservados.

El contenido de humedad en el suelo puede verse alterado por la compactación del mismo, ya que en este proceso se reducen significativamente los espacios porosos del suelo disminuyendo la disponibilidad de agua, aire y nutrientes y aumentando la resistencia de la penetración de las raíces que consecuentemente va a afectar el crecimiento y el rendimiento de las plantas, por otra parte, entre más saturado se encuentre el suelo, más vulnerable es a la compactación, especialmente los suelos arcillosos (donde las gotas de agua actúan como lubricantes y minimizan la

resistencia del suelo), es por ello que el suelo seco es menos susceptible a la compactación, al tener una mayor resistencia (Kumar et al., 2018).

2.3. Computes hierarchical clustering and cut the tree

Hierarchical Clustering (Agrupamiento Jerárquico) es un esquema de agrupamiento que construye un árbol de puntos de datos considerando la similitud entre ellos, donde la disposición de los agrupamientos formados durante el proceso se representan a través de un dendrograma, el cual se corta como resultado del agrupamiento en una medida de similitud (Silva et al., 2018; Xu et al., 2020).

El número de agrupamientos se define observando el dendrograma resultante, por lo que no es necesario que el usuario defina el número de grupos con anterioridad, este esquema suministra, de una manera más informativa, la estructura del dominio, además de que es robusto frente a los valores atípicos (Karna & Gibert, 2022).

Este método se puede clasificar de acuerdo al enfoque del agrupamiento jerárquico, siendo “aglomerativo” si los objetos de datos se agrupan ascendentemente, y “divisivos” si se agrupan de forma descendente (Karna & Gibert, 2022).

2.3.1. Silhouette method (Método de la silueta)

El *silhouette method* mide la calidad de la agrupación (Silva et al., 2018). Este método permite encontrar el número óptimo de agrupaciones, así como interpretar y validar la consistencia dentro de las agrupaciones de datos (Dudek, 2020).

El índice de silueta calcula la relación entre la dispersión máxima de la clase y la varianza para determinar la similitud de un objeto con su propio grupo (cohesión) en comparación con otro (separación) (Karna & Gibert, 2022), este se determina mediante la ecuación (1), y su valor oscila entre [+1, -1], los valores cercanos a +1 indican que el objeto posee una alta cercanía con los árboles de su misma agrupación (buena cohesión), mientras que los valores cercanos a -1 señalan una proximidad a árboles de otras agrupaciones (mala cohesión) (Dudek, 2020; Smith,

2021). Si muchos objetos tienen un valor bajo o negativo, entonces la configuración de la agrupación puede que tenga demasiadas o insuficientes agrupaciones.

$$S(u) = \frac{1}{n} \sum_{i=1}^n \frac{b(i) - a(i)}{\max\{a(i); b(i)\}} \dots (1)$$

Donde,

u = número de agrupaciones (*clusters*)

n = Número de objetos en el conjunto de datos

$\{d_{ij}\}$ = Matriz de distancia

$a(i) = \sum_{k \in \{P_r/i\}} d_{ik} / (n_r - 1)$ = Distancia media del objeto i otros objetos que pertenecen a *cluster* P_r (el objeto i pertenece al *cluster* P_r); n_r = Número de objetos en el *cluster* P_r

$b(i) = \min_{s \neq r} \{d_{iP_s}\}$, $d_{iP_s} = \sum_{k \in P_s} d_{ik} / n_s$ = Distancia media del objeto i a otros objetos que pertenecen al *cluster* P_s (el objeto i no pertenece al *cluster* P_s); n_s = Número de objetos en el *cluster* P_s

2.4. Caracterización climática del Pacífico Central de Costa Rica

La región del Pacífico Central se localiza al sureste del país, en el centro de la Vertiente del Pacífico, extendiéndose de Playa Herradura hasta Dominical, Cerros de Herradura, Cerro Turrubares, Cerro Cangreja y partes bajas de la Fila Costeña; abarcado de esta manera los poblados de Tinamaste, valle de Parrita, Quepos y Manuel Antonio (Solano Quintero & Villalobos Flores, 2001).

Su tipo de clima es lluvioso con influencia monzónica, caracterizándose por un periodo seco moderado en los meses de enero a marzo y un periodo lluvioso intenso de mayo a noviembre, donde el mes de abril y diciembre son de transición (Solano Quintero & Villalobos Flores, 2001; Instituto Meteorológico Nacional [IMN], s/f).

Debido a la proximidad del piedemonte de la Fila Costeña, los vientos del suroeste del Pacífico y la influencia de la zona de convergencia tropical, las lluvias son mayores en los valles y al sur de la región, además la protección de la Fila Costeña al este dificulta el paso directo de los vientos alisios, lo que modifica el régimen de

precipitación (Instituto Meteorológico Nacional [IMN], n.d.). La Tabla 2.4 indica la precipitación anual y la temperatura máxima y mínima en el periodo de 1961 a 1990 y de 1991 a 2005.

Tabla 2.4 Información de precipitación y temperatura en el Valle Parrita

		1961 – 1990	1991 – 2005
Precipitación anual		3558	3816
Temperatura	Máxima	30,8	31,0
	Mínima	22,6	22,8

Nota. “ El clima, su variabilidad y cambio climático en Costa Rica” por Comité Regional de Recursos Hidráulicos [CRRH], 2008, p.53 (<http://cglobal.imn.ac.cr/index.php/publications/el-clima-su-variabilidad-y-cambio-climatico-en-costa-rica/>). Derechos reservados.

Por su ubicación cercana al ecuador y su clima tropical, Costa Rica presenta una alta intensidad de radiación solar durante todo el año, siendo el promedio de radiación diaria mensual del país entre 12 y 16 MJ/m², para algunos de los cantones del Pacífico Central como Parrita y Quepos la radiación solar diaria es de 16 MJ/m² (Wright, 2009).

Las estaciones meteorológicas operan con heliógrafos (miden la duración del brillo solar) y piranómetros (miden la radiación global sobre un plano horizontal); los principales factores que afectan la radiación solar que llega a la superficie de la tierra son (Wright Gilmore, 2006):

- a. La latitud del lugar.
- b. La topografía.
- c. La cantidad nubosidad que cubre al cielo, este es el factor más importante por la alta reflectividad de las nubes.
- d. El contenido de vapor de agua y de aerosoles en la atmosfera
- e. El día del año (posición de la órbita terrestre), entre más cercana este la Tierra del Sol, ocurriendo esto en el mes de enero, mayor es la intensidad de la radiación solar, disminuyendo dicha intensidad en el mes de julio, cuando la Tierra está más alejada.

Por otra parte, debido al sistema montañoso del país que cubre más de la mitad del territorio este influye en la intensidad solar, siendo menor entre las 9 am y después de las 3 pm en territorios más bajos (Wright, 2009).

2.5. Escenario de cambio climático

El cambio climático es el cambio en el estado del clima por las variaciones en la media y en sus propiedades que continúan por un período prolongado, causado por procesos naturales internos o forzamientos externos como son las acciones antropogénicas en la composición atmosférica o en el uso del suelo (Intergovernmental Panel on Climate Change [IPCC], 2018).

El efecto más inminente del cambio climático es la elevación de la temperatura atmosférica de la Tierra (calentamiento global), ocasionado principalmente por el incremento de los gases de efecto invernadero como el dióxido de carbono (CO_2), el metano (CH_4), el ozono (O_3), el óxido nitroso (N_2O) y los clorofluorocarbonos (CFC), que absorben la radiación infrarroja y atrapan el calor en la atmósfera (Kweku et al., 2018).

De acuerdo con el IPCC (2014) (ver Figura 2.2) con el aumento de la temperatura promedio global en la superficie para el periodo 2016 – 2035 con respecto a 1986 – 2005 es similar para los cuatro escenarios RCP, en un rango de 0,3 – 0,7 °C aproximadamente, asumiendo que no se presentarían cambios por fuentes naturales. A mediados del siglo XXI la magnitud del cambio climático depende del escenario de emisiones seleccionado, mientras que a finales de siglo el aumento será de 0,3 – 1,7°C para el RCP2.6, de 1,1 – 2,6 °C según el RCP4.5, de 1,4 – 3,1°C en el RCP6.0 y de 2,6 – 4,8°C según el RCP8.5.

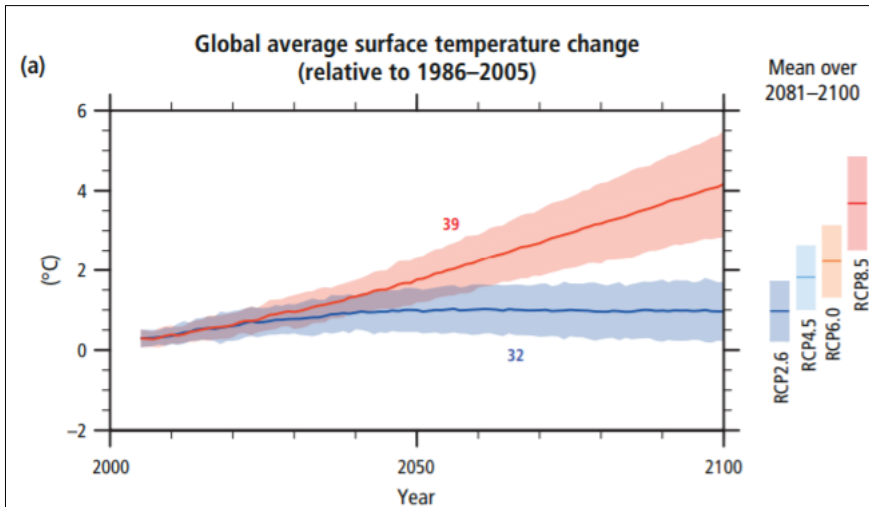


Figura 2.2 Cambio de la temperatura media global en la superficie

Nota. Todos los cambios son relativos a 1986 – 2005. Los *Representative Concentration Pathways* (RCPs) utilizados para realizar las proyecciones incluyen un escenario de mitigación estricta (RCP2.6) manteniendo el calentamiento global por debajo de los 2 °C, dos escenarios intermedios (RCP4.5 y RCP6.0) y un escenario de emisiones de GEI muy elevadas (RCP8.5). Para los escenarios RCP2.6 (azul) y RCP8.5 (rojo) se muestran las series temporales de las proyecciones y una medida de la incertidumbre (sombreado). Para todos los escenarios de RCP se presenta la media y las incertidumbres asociadas promediadas para 2081-2100 en la parte derecha del gráfico. Se indica el número de modelos del *Coupled Model Intercomparison Project Phase 5* (CMIP5) utilizados para calcular la media del multi-modelo. De *Climate change 2014: Synthesis report. Contribution of working groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change* [Imagen], por Intergovernmental Panel on Climate Change [IPCC], 2014, IPCC (<https://www.ipcc.ch/report/ar5/syr/>). Derechos reservados.

Por otra parte se pronostica a nivel global precipitaciones no uniformes según el escenario RCP8.5, donde las latitudes altas y el pacífico ecuatorial recibirían un incremento en las precipitaciones medias anuales, las regiones secas de latitudes medias y subtropicales una disminución en su precipitación anual promedio, en tanto que en regiones húmedas de latitudes medias está aumentando, además en las latitudes medias y en las regiones tropicales húmedas las precipitaciones serán posiblemente más intensas y frecuentes (IPCC, 2014).

Según Mall et al. (2017) las variaciones en la temperatura y en los patrones, así como en las precipitaciones tendrán un efecto en el contenido de agua del suelo, la escorrentía y la erosión, la salinización, la biodiversidad y el contenido de carbono orgánico y nitrógeno y la evapotranspiración; ocasionando que en regiones tropicales y subtropicales el rendimiento potencial disminuya.

A nivel país, Alvarado Gamboa (2021) llevó a cabo un estudio de cambio climático en Costa Rica para los escenarios RCP2.6 y RCP8.5 determinando temperatura, precipitación, radiación solar (ver Anexo 1 y 2), humedad relativa y velocidad del viento.

En cuanto a la región del Pacífico tanto en el escenario RCP2.6 y RCP8.5 se presenta un incremento de la temperatura, especialmente en el Pacífico Norte y Central. En el escenario RCP2.6 se expone un incremento de las precipitaciones de menor magnitud en las partes bajas del Pacífico Central y en el sur del Pacífico Sur, mientras que en las partes altas del Pacífico se da una disminución de las lluvias, a mediano plazo el aumento de las precipitaciones será entre 200 a 500 mm en las partes bajas del Pacífico Central y Sur, y una disminución de 100 y 400 mm en las zonas altas. La radiación solar será mayor en el Pacífico Norte junto con el Valle Central ($> 230 \text{ W/m}_2$). En el escenario RCP8.5, de igual manera se registra una disminución en las precipitaciones de hasta un 15% en el Pacífico Central y un 30 % en el Pacífico Sur, asimismo se presenta una subida en la región de la humedad relativa, sobre todo en el Pacífico Central y la Península de Nicoya, adicionalmente la radiación solar será mayor en el Pacífico Norte y en las zonas bajas del Pacífico Sur, y menor en el Pacífico Central y cordilleras (Alvarado Gamboa, 2021).

Hidalgo & Alfaro (2015) realizaron 107 ejecuciones climáticas de 48 modelos de circulación general (GCMs) del Coupled Model Intercomparison Project 5 (CMIP5) para evaluar y reproducir las características básicas del clima (temperatura y precipitación mensual) a finales del siglo XX en Centroamérica, siendo la mejor corrida la “cesm1_cam5(1)” con la menor sumatoria de clasificación final, lo que significa que este fue el que reprodujo de mejor manera las características de la zona de convergencia intertropical, el ciclo estacional del índice de chorro de baja intensidad del Caribe y el patrón de carga global del ENSO, esta corrida también mostró una notable habilidad para la reproducción de las desviaciones estándar de la temperatura superficial, asimismo, de forma general este modelo (es decir, “el cesm1_cam5”) tuvo un rendimiento bastante bueno en la reproducción de los

patrones medios de precipitación, y ligeramente en bueno en sus calificaciones parciales.

A partir este conjunto de datos de alta resolución (cuadrícula de 5x5 km) para Centroamérica entre 1970 a 1999, Hidalgo et al. (2017) estudiaron la variabilidad del clima generaron proyección del cambio climático mensuales para un periodo de 1979 a 2099, obteniendo para Costa Rica un pronóstico de calentamiento en gran parte del país, principalmente en la Vertiente del Pacífico Central, y un incremento de la precipitación al sur y al caribe.

2.6. WeaGETS (Weather Generator of the École de Technologie Supérieure)

Los generadores meteorológicos se fundamentan en la cadena de Markov de primer orden y en la distribución gamma, permitiendo simular la precipitación diaria, la temperatura y la radiación solar manteniendo sus estadísticas a largo plazo, lo cual los convierte en una herramienta clave para estudios hidrológicos, agrícolas, de reducción de escala y de impacto climático (Tseng, Chen, & Senarath, 2020).

Weather Generator École de Technologie Supérieure (WeaGETS) es un generador meteorológico estocástico basado en MatLab, este produce series diarias de precipitación y temperatura (máximas y mínimas) de longitud ilimitada, lo que posibilita ejecutar estudios de impacto de variables meteorológicas de baja frecuencia y de cambio climático, asimismo, este generador se caracteriza por introducir esquemas computacionales de otros generadores meteorológicos (Chen et al., 2012; Herrera et al., 2017).

WeaGETS permite generar la ocurrencia de la precipitación aplicando el modelo de cadena de Markov de primer, segundo y tercer orden, siendo la de primer orden la más sencilla y utilizada, en ella se define un día húmedo o seco de acuerdo a la probabilidad de que se produzca precipitación, y que depende del estado del día anterior (Yang et al., 2022). Esta probabilidad de transferencia de primer orden (P_{01} y P_{11}) se expresan de la siguiente manera:

$$P_{01} = P\{\text{Precipitación en el día } t \mid \text{Ninguna precipitación en el día } t - 1\} \dots (2)$$

$$P_{11} = P\{\text{Precipitación en el día } t \mid \text{Precipitación en el día } t - 1\} \dots (3)$$

Donde,

P_{01} = Probabilidad de precipitación en el día t en caso de que no haya precipitación en el día $t - 1$

P_{11} = Probabilidad de precipitación tanto en el día t como en el día $t - 1$

Por lo que al existir dos casos de eventos de precipitación en un día determinado, las probabilidades de transición de las dos probabilidad de transferencia son $P_{00} = 1 - P_{01}$ y

$P_{10} = 1 - P_{11}$ (Yang et al., 2022). Estas fórmulas pueden generalizarse para cadenas de Markov de segundo y tercer grado como son las siguientes:

$$P_{ijk} = P\{R_t = k | R_t = j | R_t = i\} \dots (4)$$

$$P_{hijk} = P\{R_t = k | R_t = j | R_t = i | R_t = h\} \dots (5)$$

Donde,

$h, i, j, k = 0$ o 1

$R_t = 1$, si la precipitación ocurre en el día t , es decir, el día t es un día húmedo

$R_t = 0$, si no ocurre precipitación en el día t , es decir, el día t es un día seco

Para obtener la cantidad de precipitación de un día húmedo se dispone de dos funciones de distribución de probabilidad, la primera es una distribución exponencial de un parámetro cuya función de densidad de probabilidad está dada por la ecuación 6; y la segunda es una distribución gamma de dos parámetros, dada por la ecuación 7 (Yang et al., 2022).

$$f(x) = \lambda e^{-\lambda x} \dots (6)$$

$$f(x) = \frac{\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta \Gamma(\alpha)} \dots (7)$$

Donde,

λ = Parámetro de distribución exponencial (igual a la inversa de la media numérica)

x = Intensidad de precipitación diaria

β y α = Parámetro de la distribución gamma, de escala y forma, respectivamente

$\Gamma(\alpha)$ = Valor de la función gamma evaluada en α

Debido a la facilidad de cálculo y sus resultados, la distribución gamma es la más utilizada para generar la cantidad de lluvia (Chen et al., 2012).

En lo que respecta a la generación de temperatura máxima y mínima se utiliza un modelo lineal autorregresivo de primer orden, este consiste en reducir a elementos residuales la serie temporal de datos observados como la diferencia entre las medias diarias y dividiéndolas por las desviaciones estándar, las cuales están condicionados por el estado húmedo o seco (ver ecuación 8)

$$\chi_{p,i}(j) = A\chi_{p,i-1}(j) + B\varepsilon_{p,i}(j) \dots (8)$$

Donde,

$\chi_{p,i}(j)$ = Matriz (2x1) para el día i del año p , cuyos elementos son los residuos de T_{\max} ($j = 1$) y T_{\min} ($j = 2$)

$\varepsilon_{p,i}(j)$ = Matriz (2x1) de componentes aleatorios independientes que se distribuyen normalmente con una media de cero y una varianza de la unidad

$A = M_1 M_0^{-1}$ y $B B^T = M_0 - M_1 M_0^{-1} M_1^T$ = Matriz (2x2) cuyos elementos están definidos de manera que las nuevas secuencias tienen los coeficientes de autocorrelación y correlación cruzada deseados; -1 y T = Inversa y transposición de la matriz, respectivamente; M_0 y M_1 = Matrices de covarianza del retardo 0 y del retardo 1, respectivamente.

Para generar la temperatura máxima y mínima se utiliza el esquema condicionado en función de la desviación estándar, si la desviación estándar de T_{\max} es mayor o igual que la de T_{\min} las temperaturas se obtienen por medio de las ecuaciones 9 y 10, caso contrario, se obtienen con las ecuaciones 11 y 12 (Chen et al., 2012).

$$T_{\min} = \mu_{\min} + \sigma_{\min} \times \chi_{p,i} \dots (9)$$

$$T_{max} = T_{min} + (\mu_{max} - \mu_{min}) + \sqrt{\sigma_{max}^2 - \sigma_{min}^2} \times \chi_{p,i} \dots (10)$$

$$T_{max} = \mu_{max} + \sigma_{max} \times \chi_{p,i} \dots (11)$$

$$T_{min} = T_{max} - (\mu_{max} - \mu_{min}) - \sqrt{\sigma_{min}^2 - \sigma_{max}^2} \times \chi_{p,i} \dots (12)$$

Donde,

μ = Media mensual

σ = Desviación estándar

χ = Desviación normal generada

En WeaGETS es posible eliminar las variaciones bruscas de los parámetros y suavizar la distribución calculada de los parámetros diarios, permitiendo reproducir las características de precipitación de la línea suavizada (variando ligeramente con la precipitación observada) mediante cuatro armónicos de Fourier; por otro lado por medio de un enfoque de corrección espectral WeaGETS ajusta la subestimación de la varianza mensual e interanual de la precipitación y la temperatura (J. Chen, F. P. Brissette, R. Leconte, & A. Caron, 2012).

Lo descrito anteriormente acerca de la ejecución del generador meteorológico WeaGETS se resumen en la Figura 2.3.

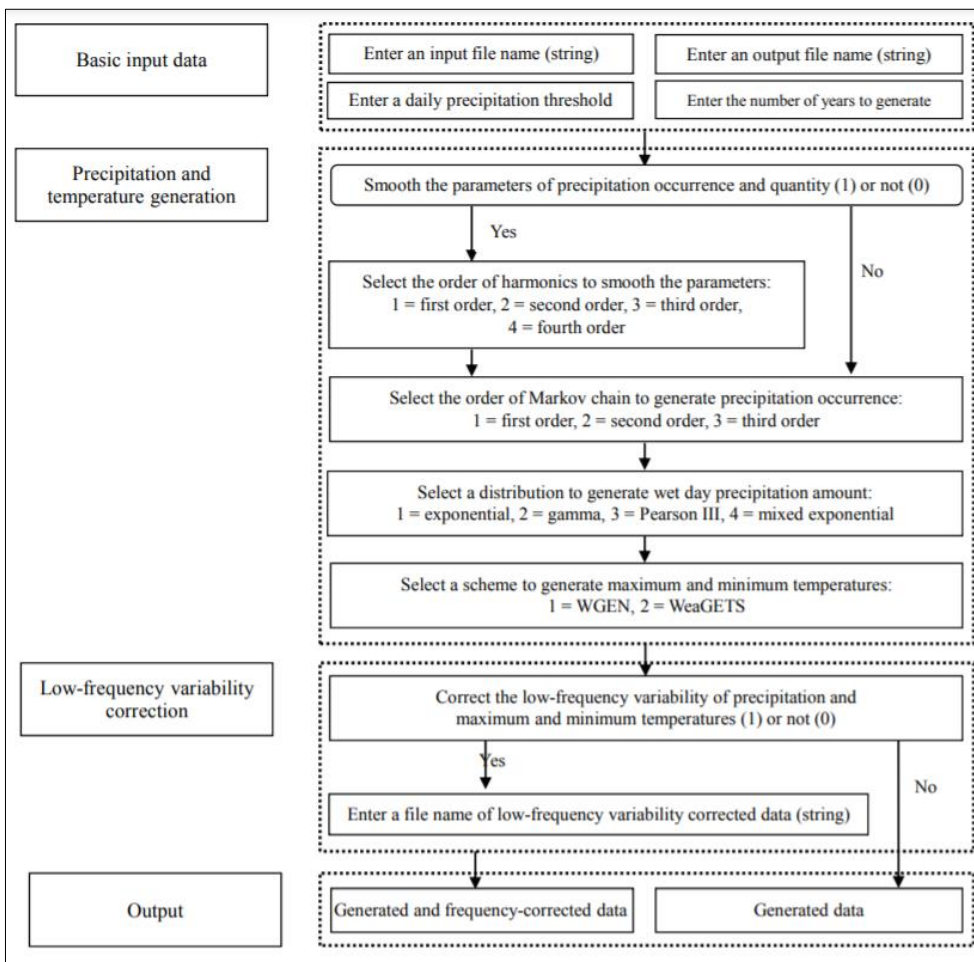


Figura 2.3 Estructura del generador meteorológico estocástico WeaGETS

Nota. A Versatile Weather Generator for Daily Precipitation and Temperature [Imagen] por J. Chen, F. P. Brissette, R. Leconte, & A. Caron, 2012, *Transactions of the ASABE*, 55(3), 895–906. (<https://doi.org/10.13031/2013.41522>). Derechos reservados.

2.7. Modelos de aprendizaje automático

2.7.1. Random Forest (RF)

Random Forest (RF) es un algoritmo de aprendizaje basado en conjuntos formado por árboles de decisión en el que cada árbol depende de una colección de variables aleatorias (Cutler, Cutler, & Stevens, 2012). Los árboles de decisión se construyen al dividir recursivamente en dos grupos la serie de datos dada, esto según un criterio determinado hasta que se cumpla una condición de parada preestablecida, de acuerdo en cómo se definan estos criterios (partición y parada) los árboles de decisión pueden ser para funciones de clasificación o regresión, dichos criterios se plantean como un problema de optimización (Schonlau & Zou, 2020). El modelo RF se caracteriza por su precisión en la predicción, la generalización del error y la velocidad de convergencia (Huo, Shi, & Chang, 2016).

RF en regresión es un método de aprendizaje estadístico que utiliza la técnica de agregación *Bootstrap* para obtener múltiples muestras de la información original, para posteriormente reunir los árboles de decisión formados, tomando la predicción promedio de los árboles de decisión para generar los resultados y realizar el proceso de estimación (Huo et al., 2016). Para tareas de regresión generalmente se utiliza el error cuadrático medio como criterio de división en cada nodo interno (Schonlau & Zou, 2020).

2.7.2. Neural Network (NN)

Neural Network (NN) realiza una operación matemática mediante un modelo determinista (función no lineal) con el objetivo de calcular las salidas a partir de un conjunto de datos de entrada; en virtud de su función la red neuronal se divide en una capa de entrada, una capa oculta (que es la capa de procesamiento interno de la red) y una capa de salida, las cuales están conectadas; cada neurona o nodo se activa si la cantidad total de señal recibida supera cierto umbral de activación (Ciaburro & Venkateswaran, 2017; Lee et al., 2017).

NN es una excelente herramienta estadística y de predicción, por su función flexible para entrenar la red neuronal (esto es, para aproximar una relación funcional en la situación anterior), además de que no necesidad de definir con anterioridad el tipo relación entre las covariables y las variables de respuesta (Günther & Fritsch, 2010). Por otro lado, el rendimiento y optimización de este modelo se ve afectado por el número de capas ocultas, el número de nodos dentro de la capa oculta, el número de iteraciones de aprendizaje, la tasa de aprendizaje y el impulso (Lee et al., 2017).

Los parámetros de configuración de los paquetes en R de estos modelos se muestran en la Tabla 2.5.

Tabla 2.5 Parámetros utilizados en la configuración de los modelos propuestos

Modelo	Parámetro	Descripción
RF	ntree	Número de árboles

	mtry	Número de variables muestreadas aleatoriamente como candidatas en cada división
NN	threshold	Valor numérico que especifica el umbral de las derivadas parciales de la función de error como criterio de parada
	stepmax	Pasos máximos para el entrenamiento de la red neuronal
	algorithm	Cadena que contiene el tipo de algoritmo para calcular la red neuronal

Nota. Adaptado de " Comparison of Random Forest and SVM for electrical short-term load forecast with different data sources" por J. Huo, T. Shi, & J. Chang, 2016, del 26 al 28 de agosto, 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS) [Sesión de conferencia], Beijing, China (<https://doi.org/10.1109/ICSESS.2016.7883252>) y "neuralnet: Training of Neural Networks" por F. Günther, & S. Fritsch, 2010, *The R Journal*, 2(1), 30–38 (<https://journal.r-project.org/archive/2010/RJ-2010-006/index.html>). Derechos reservados.

2.8. Coeficientes de medición del error y deficiencia

Cabe agregar que los modelos deben calibrarse para obtener resultados confiables utilizando uno o varios índices con el fin de evaluar detenidamente el resultado de la simulación del modelo (Lin, Chen, & Yao, 2017). Entre ellos destacan:

- a. Coeficiente de determinación (R^2): Describe la proporción de la varianza de los datos medidos explicada por el modelo, sus valor oscila entre 0 y 1, donde los valores más altos indican una menor varianza del error (Moriasi et al., 2007). La precisión del modelo se puede clasificar como insatisfactorio con R^2 menores a 0,6, satisfactorios de 0,6 a 0,7, buenos de 0,7 a 0,8 y muy buenos si son mayores a 0,8 (Bosi, Sentelhas, Pezzopane, & Santos, 2020).
- b. Coeficiente de eficiencia de Nash-Sutcliffe (NSE): Refleja de mejor manera las características deseables e indeseables del modelo de interés, entre mejor sea la calidad del modelo mayor es el coeficiente, es decir, la calidad perfecta deseable (Zeybek, 2018), su rango oscila entre $-\infty$ y 1, siendo valores menores a 0,5 insatisfactorios, de 0,5 a 0,65 satisfactorios, de 0,65 a 0,75 buenos, y muy buenos de 0,75 a 1; aunque si el valor del NSE y el R^2 es mayor a 0,5 y 0,6 respectivamente, el resultado del modelo es aceptable (Moriasi et al., 2007).
- c. Error absoluto medio (MAE): Indica la diferencia entre los valores observados y los estimados por el modelo (Gomes, Bosi, Pedreira, Santos, & Pedreira, 2020), proporcionando la magnitud absoluta de la estimación del modelo (Neil I. Huth, Banabas, Nelson, & Webb, 2014).
- d. Error cuadrático medio (RMSE): Medida de la desviación entre el resultado de la estimación del modelo comparado con el valor efectivo, es decir, el observado (Gomes et al., 2020). Un RMSE pequeño indica una desviación menor y por ende un mejor rendimiento del modelo (Yanyang, Xu, Zhao, Li, & He, 2017).

2.9. Modelo Agricultural Production Systems sIMulator (APSIM)

Agricultural Production Systems sIMulator (APSIM) es una plataforma para la modelización y simulación de procesos biofísicos en sistemas agrícolas desarrollada por la *Agricultural Production Systems Research Unit* en Australia, es una de las principales herramientas para conocer los resultados económicos y ambientales de la gestión agrícola ante riesgos climáticos (Keating et al., 2003). La Tabla 2.6 indica el uso del modelo APSIM en estudios a nivel mundial en distintas aplicaciones.

Tabla 2.6 Uso de APSIM a nivel mundial

Aplicación	País de estudio	Artículo	Descripción
Investigación de procesos de sostenibilidad del suelo agrícola	Australia	(Luo, Wang, Baldock, & Xing, 2014)	Simulación del carbono orgánico del suelo a través del tiempo y el espacio bajo 59 sistemas de cultivos de cereales.
Modelización de sistemas de cultivo hortícola	Australia	(Huth et al., 2009)	Aplicación de APSIM-Broccoli para describir las respuestas observadas del cultivo en el desarrollo del dosel y la iniciación floral.
Modelización de sistemas agroforestales	Papúa Nueva Guinea	(Neil I. Huth et al., 2014)	Desarrollo de un modelo de producción para palma de aceite, entregando las principales respuestas del crecimiento al clima y a la gestión.
Evaluación del uso y la eficiencia de los recursos	Australia	(Hunt et al., 2013)	Uso del modelo APSIM para complementar experimentos de campo, simulando el efecto del agua y el N en suelos arenosos y arcillosos en barbecho.
Caracterización ambiental	Australia	(Chauhan, Solomon, & Rodriguez, 2013)	Simulación de la producción de maíz en distintas regiones, configurando el modelo de acuerdo con la relación diaria de demanda de agua como indicador del grado de estrés hídrico y el rendimiento final.
Evaluaciones de la brecha de rendimiento	Australia, China y Zimbabue	(Carberry et al., 2013)	Simulación del rendimiento de granos para diagnosticar el estado de la producción agrícola y sugerir vías para mejorar la eficiencia y reducir la brecha de rendimiento.
	Australia	(Hochman, Gobbett, Horan, & Navarro Garcia, 2016)	Determinación de la brecha del rendimiento del trigo, entre el rendimiento real y el rendimiento simulado con limitación hídrica.
Información sobre programas de fitomejoramiento	Estados Unidos	(Messina, Podlich, Dong, Samples, & Cooper, 2011)	Simulación de fenotipos con APSIM-maize (variación del dosel arbóreo, el ángulo de la raíz, el crecimiento de la espiga, intercepción de la luz, la absorción de agua, la asignación de biomasa, entre otros).
Modificación hipotética de rasgos	Australia	(Lilley & Kirkegaard, 2016)	Simulación de cultivo de trigo a largo plazo en un entorno semiárido con modificación de su sistema radicular.
Análisis del cambio climático y de la adaptación	Australia	(Wang, Wang, & Liu, 2011)	Simulaciones del cultivo de trigo para clima de referencia y 14 simulaciones con cambios de temperatura, niveles de CO ₂ y precipitación.

Investigación de las limitaciones bióticas y abióticas del sistema	Australia	(Whish, Herrmann, White, Moore, & Kriticos, 2015)	Diseño de un método que permita que los modelos de población de roya en las hojas del cultivo en DYMEX interactúen con APSIM, modificando el crecimiento y desarrollo del área foliar.
--	-----------	---	--

Nota. Adaptado de "APSIM Next Generation: Overcoming challenges in modernising a farming systems model" por D. Holzworth, N.I. Huth, J. Fainges, H. Brown, E. Zurcher, R. Cichota, S. Verrall, N.I. Herrmann, B. Zheng, & V. Snow, 2018, *Environmental Modelling & Software*, 103, 43–51. (<https://doi.org/10.1016/j.envsoft.2018.02.002>). Derechos reservados.

APSIM permite conectar a su motor diferentes componentes de un sistema agrícola, la Figura 2.4 muestra diagrama de los módulos utilizados en este estudio, los cuales son una unidad de cálculo y representan una colección de procesos (D. P. Holzworth et al., 2014), entre ellos se encuentran el módulo de cultivo de palma de aceite, de suelo, de prácticas de gestión y meteorológico.

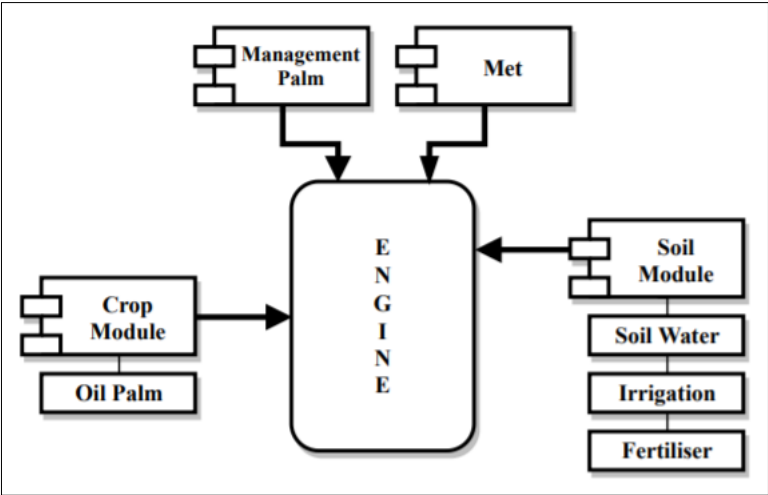


Figura 2.4 Diagrama de flujo de los diferentes módulos del modelo APSIM

El módulo de cultivo (*Crop Module*) simula el crecimiento de una especie a través del tiempo en respuesta a los aportes del clima, del suelo y las prácticas de manejo (APSIM Initiative [AI], n.d.); un esquema de su funcionamiento se muestra en la Figura 2.5.

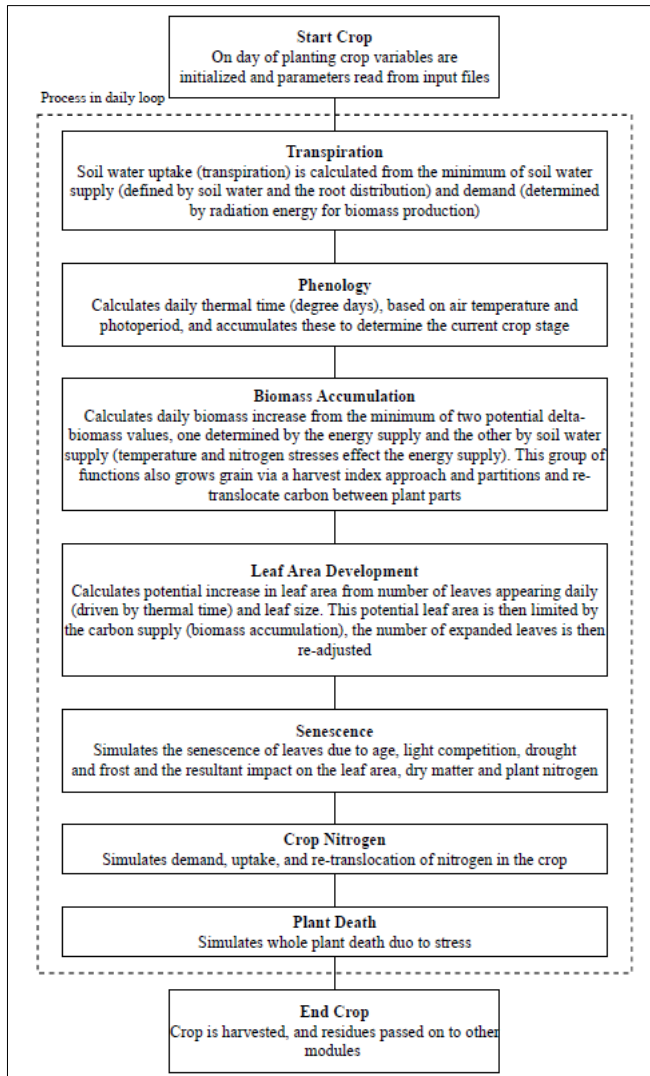


Figura 2.5 Esquema de Crop Module

Nota. De *Documentation: APSIM 7.10*. [Imagén] por APSIM Initiative [AI], s/f (<https://www.apsim.info/documentation/model-documentation/>). Derechos reservados

Huth et al. (2014) desarrollaron y validaron el sistema de producción para palma de aceite en APSIM, comparando el crecimiento real observado y el de la biomasa modelada, durante un periodo de 17 años. *APSIM-Oil Palm* simula el crecimiento de las frondas, el tallo, las raíces y los racimos de la palma, además de calcular el crecimiento y el desarrollo de la planta, también estima el uso de recursos y los flujos de materia orgánica, compartiendo esta información a otros modelos dentro de la simulación (Neil I. Huth et al., 2014); asimismo incorpora el desarrollo del cultivo del sotobosque, que muere a medida que la plantación madura (Akram, Levia, Herrick, Lydiasari, & Schütze, 2022).

El módulo meteorológico (*Met Module*) de APSIM requiere de datos meteorológicos diarios para especificar el clima del sitio en estudio, estos corresponden a la precipitación, la temperatura máxima y mínima y la radiación solar (Okoro, Schickhoff, Boehner, Schneider, & Huth, 2017) (AI, s/f).

APSIM-*Soil* simula los procesos que ocurren sobre y en el perfil del suelo, como es la infiltración y el movimiento del agua, la evaporación, la escorrentía y el drenaje, descomposición de materia orgánica, entre otros (D. P. Holzworth et al., 2014); las entradas básicas de para el crecimiento del cultivo en el suelo incluyen parámetros físicos y/o químicos (Vogeler et al., 2022), en el caso de este estudio corresponden a la densidad aparente, la capacidad de campo (*Drained upper limit* - DULL), el punto de marchitez permanente (*15Bar lower limit* - LL15), la capacidad de saturación (*Soil water saturation* - SAT) y el suelo seco (*Soil air dry* - AirDry). Para simular el movimiento del agua en el perfil del suelo *SoilWater* es un modelo de balance hídrico en cascada basado en los algoritmos de la familia de modelos CERES, donde las características hídricas del suelo se especifican en términos de los contenidos volumétricos del punto de marchitez permanente (*15Bar lower limit* - LL15), de la capacidad de campo (*Drained upper limit* - DULL), y de la capacidad de saturación (*Soil water saturation* - SAT) (AI, s/f). Según Akram et al. (2022) *SoilWater* puede dar errores de predicción sobre el movimiento del agua en suelos saturados, puesto que fue desarrollado para suelos no saturados, además de tener cierta imprecisión al predecir los efectos del movimiento lateral del agua del suelo dado que es un modelo de cubeta en cascada.

Mientras que para aplicación de fertilizantes sólidos al sistema existe el módulo de fertilizantes *Fertiliser*, mediante el cual calendario de varios años de aplicación, donde el usuario requiere especificar la cantidad y el tipo de fertilizante y la profundidad de aplicación (AI, s/f).

El módulo de riego de APSIM (*Irrigation*) permite al usuario especificar programas de riego de varios años y/o configurar un programa de riego automatizado en función de la humedad del suelo, así como activar o desactivar dichos programar cuando el usuario lo quiera, también se pueden aplicar solutos en el agua de riego (AI, s/f).

Algunos estudios más recientes del uso de APSIM en palma de aceite son los siguientes:

Culman et al. (2019) desarrollaron un método para evitar que la palma fuera sometida a estrés por déficit hídrico, utilizando una técnica de fusión de datos y la evaluación del método en el rendimiento del cultivo mediante dos modelos de simulación de cultivos, uno de ellos siendo APSIM, en el cual se simularon dos tratamientos: con riego manejado por el método de inferencia y sin riego. Para ellos las principales variables para la aplicación del riego en la palma fueron la humedad del suelo (promedio cada 3 días) y la presión de vapor (máxima de 1,8 kPa), los resultados de la simulación revelaron un aumento del 27% en dos años de la producción de racimos con aplicación de riego.

Akram et al. (2022) indica que la demanda de agua adicional en las plantaciones de palma de aceite no solo se requiere en zonas de baja precipitación, también en aquellas zonas con características edáficas diferentes. En su estudio se empleó

APSIM para simular el crecimiento de la palma de aceite en tierras marginales (suelo pirítico) durante un periodo de 8 años y poder estimar las necesidades de riego en las distintas fases de crecimiento de la planta, llevando a cabo dos escenarios: con y sin riego, donde la aplicación del agua de riego inicialmente se planteó en función de la lluvia, si esta era menor a 20 mm en tres días consecutivos se aplicaban 25 mm de agua al día con una eficiencia del sistema del 80%. Obtuvieron que las necesidades de agua aumentan considerablemente durante los 4 y 5 años, además de que debido a que el tipo de suelo precisa un alto nivel freático se requiere de aproximadamente 80 a 95% más de agua de riego.

Por otro lado, Okoro et al. (2017) evaluaron el impacto del cambio climático de diferentes modelos de circulación general (GCMs) sobre el rendimiento de la palma de aceite en el Delta del Níger utilizando el modelo de simulación APSIM, obteniendo principalmente que el impacto neto por cambio climático es positivo y dinámicamente inconsistente a lo largo del intervalo de las simulaciones, que las diferencias de rendimientos entre los escenarios de RCP (4.5 y 8.5) fueron pequeñas y que los rendimientos de la palma son más resistentes a un incremento de las precipitaciones que de temperatura. Asimismo, recomiendan cambiar las fechas de siembra y la aplicación de riego para optimizar el rendimiento del cultivo bajo un marco de cambio climático.

3. METODOLOGÍA

3.1. Selección de lotes de trabajo

El análisis del rendimiento de cosecha se realizó utilizando la herramienta Rstudio. La base de datos se creó contemplando columnas para fecha, edad del cultivo, año de siembra, rendimiento (t/ha) y lote. La serie de datos tienen una escala mensual y comprende desde 1994 hasta el 2015.

La selección de los lotes de trabajo se subdividió en dos partes. Una preselección que consistió en filtrar los lotes que tuvieran como fechas de siembra de 1995 a 1996 y que contaran con una cantidad de datos faltantes menor al 5%.

La segunda parte consistió en realizar un agrupamiento de los lotes preseleccionados según el comportamiento del rendimiento. Definiendo el número de grupos óptimos a través del método de Silhouette Method. Para esto se utilizó la función `fviz_nbclust` del paquete `factoextra` y la realización del clúster se hizo con el método de `Computes Hierarchical Clustering and Cut the Tree` aplicando la herramienta `hcut` del mismo paquete mencionado anteriormente. Una vez realizado el agrupamiento y definido el corte de cada grupo, se seleccionó de forma aleatoria un lote que representara a cada de ellos.

3.2. Parámetros físicos del suelo

En cada lote seleccionado se seleccionaron dos puntos de muestreo. La cantidad de puntos utilizada dependió de aspectos operativos y presupuestarios, en donde se contempló el tiempo de recolección, costo de equipo y laboratorio, además, del tiempo de las pruebas físicas del suelo. Las pruebas realizadas son determinación de textura, curvas de humedad y densidad aparente.

La recolección en cada punto se realiza a lo largo de un perfil de suelo de 90 cm de profundidad, tomando entre 4 y 5 muestras repartidas a lo largo del perfil. Las pruebas realizadas son determinación de textura, curvas de humedad y densidad aparente.

3.3. Análisis de las series de tiempo de datos climatológicos de precipitación, temperatura y radiación

Los datos climatológicos utilizados son precipitación diaria (P), radiación (Rad), temperatura máxima (maxT) y mínima (minT). Estos datos son de 10 estaciones pluviométricas pertenecientes a la empresa Palma Tica S.A. Los datos de temperatura y radiación fueron suministrados por el Instituto Meteorológico Nacional de las estaciones Damas-Quepos y Finca Damas. El periodo de tiempo de las variables meteorológicas observadas mediante instrumentos en sitio va desde 1995 hasta el 2015.

El análisis conllevó, en primera instancia, la determinación de los porcentajes de datos faltantes, mismos que fueron rellanados utilizando regresiones multivariable:

$$i = (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni}) + e_i \dots (13)$$

Donde,

Y_i = Variable dependiente (Estación a rellenar)

X_i = Variable predictora (Estaciones con mayor correlación)

i = Efecto promedio que tiene el incremento en una unidad de la variable predictora X_i sobre la variable dependiente. Se conocen como coeficientes parciales de regresión

e_i = Residuo o error, la diferencia entre el valor observado y el estimado por el modelo

Completadas las series de datos meteorológicos, aprovechando que se tiene una red robusta de estaciones pluviométricas, se interpoló a cada uno de los centroides de los lotes seleccionados para contemplar la diferencia espacial en el comportamiento de la lluvia en los modelos de producción agrícola. La interpolación se realizó utilizando la metodología IDW (ver ecuación 14).

$$z_p = \frac{\sum_{i=1}^n \left(\frac{z_i}{d_i^p} \right)}{\sum_{i=1}^n \left(\frac{1}{d_i^p} \right)} \dots (14)$$

Donde, z_p es el valor estimado en el punto p , d_i^p es distancia desde el punto i al punto p , z_i es el valor conocido en el i –ésimo punto, p *exponente* es la potencia del inverso de la distancia.

3.4. Escenario de cambio climático

El escenario de cambio climático (ECC) que se utilizó como base del estudio fue generado por Hidalgo et al (2015) para el periodo de 1979 al 2099. Específicamente, se utilizará el escenario cesm1_cam5 (modelo del Centro Euromediterráneo para el Cambio Climático), debido a que fue la simulación que mejor capacidad tuvo para reproducir las características básicas de las variables precipitación y temperatura mensual de un total de 107 corridas.

Por tener una escala temporal a nivel mensual se realizó un procedimiento de disgregación para transformar los datos de acumulados mensuales a acumulados diarios. El proceso consta de 3 pasos:

- i. El primer paso, consiste en utilizar la herramienta Stochastic Weather Generator (WeaGETS) para generar 5000 años de datos diarios de precipitación (P), temperatura máxima (maxT) y mínima (minT). Para lo cual la herramienta requiere los datos observados de las variables P, maxT y minT con un periodo comprendido entre 1995 y 2015.
- ii. En segundo paso se realiza una selección cíclica, mes a mes de los datos estocásticos, de forma que coincidan con el escenario de cambio climático. Por ejemplo: se toma el dato acumulado de precipitación mensual de enero de 1995 del ECC (valor de búsqueda) y se explora dentro de los datos estocásticos los 5000 meses de enero generados y a cada uno se le obtiene el acumulado mensual y se selecciona el que más se aproxime al valor de búsqueda. Esto se realizó para toda la serie de datos mensuales del ECC, con lo que se obtuvo una serie de datos diarios de P, maxT y minT de ECC cuyos valores mensuales acumulados y promedios coinciden. Para realizar este proceso se creó un algoritmo que ejecuta el proceso de forma automática.
- iii. Finalmente se realizó un ajuste de media y desviación estándar de los datos diarios del ECC diario en función de datos observados y respetando la variabilidad interanual de las variables. Obteniendo una serie de tiempo diaria de datos de P, maxT y minT.

Para generar la serie de radiación correspondiente al escenario de cambio climático, se evaluaron dos modelos de aprendizaje automático, un Neural Network (NN) y un Random Forest (RF). Estos modelos se construyeron a partir de los datos de Rad, P, maxT y minT obtenidos con estaciones y se adicionó el número de día del año como una variable más. El proceso de construcción de los modelos se describe a continuación:

- Los datos se dividen en etapa de entrenamiento (Training Data) y validación (Test Data) de manera aleatoria distribuidos con una relación 70% y 30% de los datos respectivamente.

- En el proceso de entrenamiento, se calibran los modelos y se evalúan por medio de los coeficientes, de eficiencia de Nash-Sutcliffe (NSE), la raíz del error cuadrático medio (RMSE), el coeficiente medio de error absoluto (MAE) y el coeficiente de determinación (R^2) (ver ecuaciones 15 – 18)

$$NSE = 1 - \frac{\sum_{i=1}^n (y_i^{obs} - y_i^{sim})^2}{\sum_{i=1}^n (y_i^{obs} - \bar{y}^{obs})^2} \dots (15)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i^{obs} - y_i^{sim})^2} \dots (16)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i^{obs} - y_i^{sim}| \dots (17)$$

$$R^2 = \left(\frac{\sum [(y_i^{obs} - \bar{y}^{obs})(y_i^{sim} - \bar{y}^{sim})]}{\sqrt{\sum (y_i^{obs} - \bar{y}^{obs})^2 * \sum (y_i^{sim} - \bar{y}^{sim})^2}} \right)^2 \dots (18)$$

Donde, n es la cantidad de datos disponibles para el estudio, y_i^{obs} corresponde a la información de la edad del cultivo obtenido para el intervalo en estudio, \bar{y}^{obs} es el promedio de la edad del cultivo obtenido para el intervalo en estudio, y_i^{sim} es la edad del cultivo simulada y \bar{y}^{sim} es el promedio de la edad del cultivo simulada.

- En la validación se utiliza las series de datos Test Data en los modelos de (RF) y Neural Network (NN), y son evaluados nuevamente con NSE, RMSE, MAE y R^2 . Seleccionado el modelo que mejor calificado resulte en este proceso de validación.
- Finalmente, el mejor modelo se utilizó para generar la serie de datos de Radiación a partir de los datos de P, maxT y minT del escenario de cambio climático.

3.5. Simulación de la producción agrícola utilizando el modelo APSIM

En el montaje del modelo APSIM, se utilizaron los siguientes módulos; meteorológico, OilPalm, fertilización, materia orgánica superficial, Mantenimiento, Suelo y el de irrigación que solamente se utilizó para determinar las láminas de riego requeridas para atenuar los efectos producidos bajo los diversos escenarios de cambio climático.

Básicamente, son cuatro fases de modelación con APSIM:

- i. Proceso de calibración en donde se generan modelos para cada uno de los lotes seleccionados, las funciones de objetivo utilizadas fueron $NSE > 0.6$, $r^2 > 0.65$, según los valores obtenidos por los creadores del módulo para palma de aceite (Huth et

al., 2014). Esta calibración se realiza por medio del ajuste de valores de humedad en suelo específicamente AirDry, LL15, DULL y SAT.

- ii. Con los modelos calibrados para cada lote, se realizan tres tipos de simulación, estos serían; línea base entre los años 2000 y 2019 (Base), un primer escenario de cambio climático del 2040 al 2059 (ECC1) y un segundo escenario de fin de siglo del 2080 al 2099 (ECC2). Con el objeto de definir el impacto del cambio climático sobre la producción. Lo cual se logra al comparar el rendimiento de la línea base contra los dos escenarios de cambio climático, y así determinado si existe o no una disminución en el rendimiento por edad de cultivo y en el rendimiento promedio a lo largo de los 20 años de simulación.
- iii. En esta tercera fase, lo que se realiza es el mismo procedimiento del punto anterior, pero activando el módulo de riego, para determinar cuál sería la lámina neta requerida para mantener los rendimientos obtenidos en la línea base (2000 – 2019).
- iv. Finalmente se cuantificó los volúmenes totales de agua requeridos y se analizó la factibilidad hídrica.

En la Figura 3.1 se muestra de forma esquemática la metodología utilizada en el presente estudio.

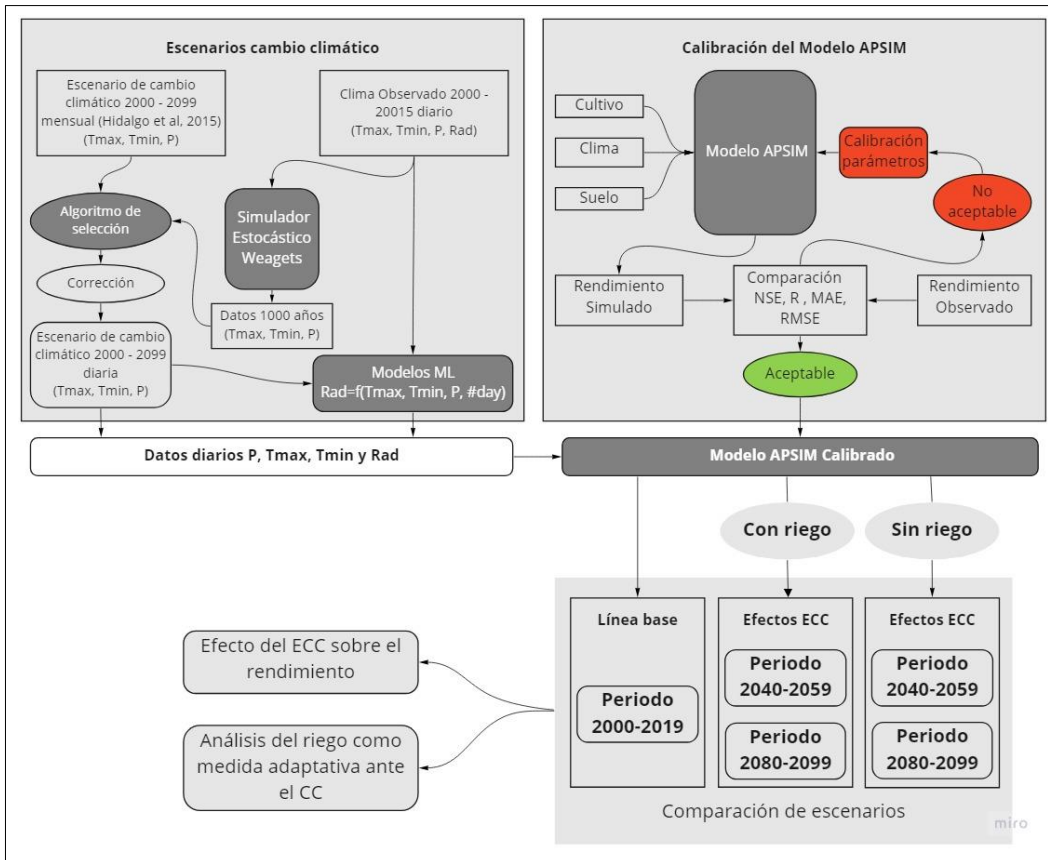


Figura 3.1 Diagrama metodológico del análisis de la información

4. RESULTADOS

El proyecto se realiza en la plantación de palma de aceite aprovechada por la compañía Palma Tica S.A, ubicada en el Pacífico Central de Costa Rica. El área de siembra es de 9445.74ha subdivididas en un total de 186 Lotes (Figura 4.1). La precipitación anual promedio de la zona es de 4537 mm y su rango promedio de temperaturas es de 22°C a 32°C. Comprende una época húmeda que se extiende del mes de mayo hasta noviembre y una época seca comprendida entre diciembre y abril.

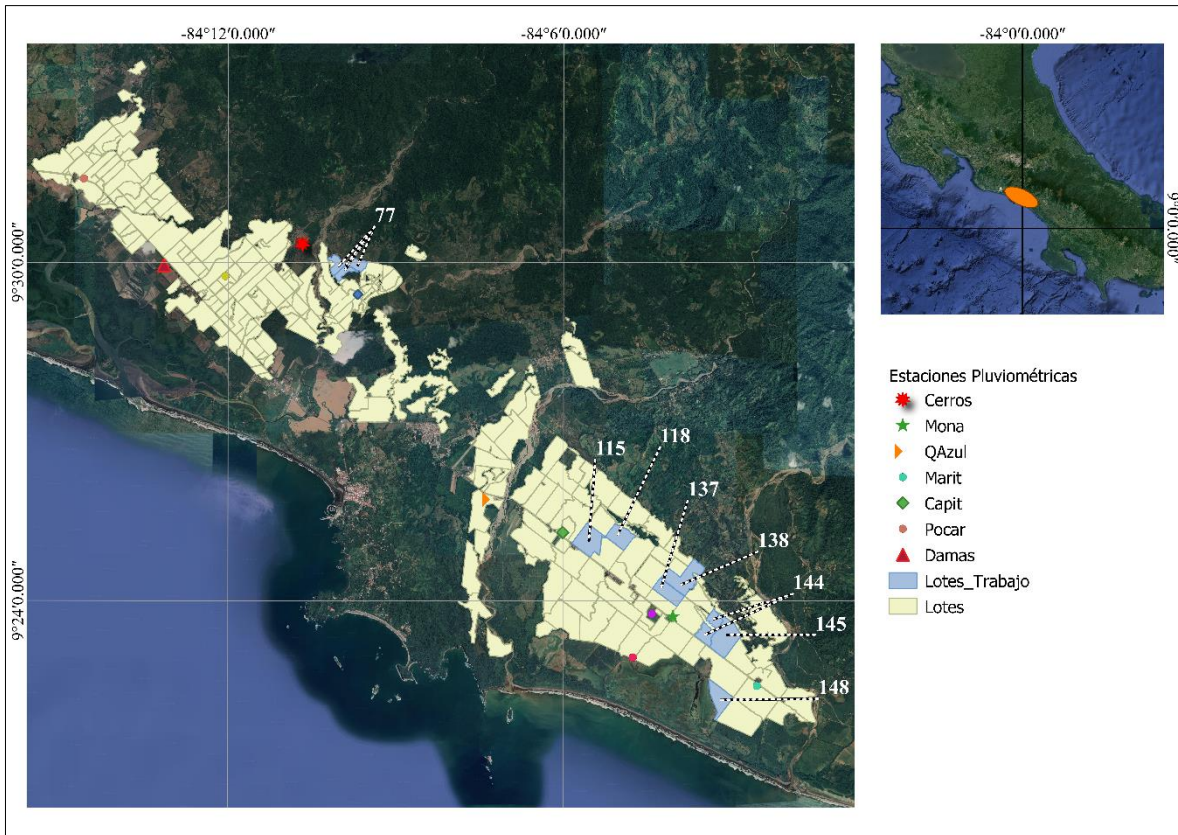


Figura 4.1 Ubicación de la plantación de la compañía Palma Tica.S.A

En general, analizados los 8 puntos de muestreo, se determinó que la textura más predominante es la franca arcillosa, seguida de texturas más pesadas hasta llegar a suelos arcillosos. Si se observan las muestras por separadas por profundidad de muestreo, no se observa una tendencia homogénea de cambio en la textura, según se muestra en la figura 4.5.

4.1. Selección de lotes de trabajo

El número de lotes a utilizar en el estudio está en función del número de unidades de producción homogéneas o clústeres generados. Para determinar el número de lotes seleccionados que posteriormente se utilizaron en el modelo APSIM, utilizando el método de Silhouette (Figura 4.2) se determinó que el número de agrupamientos mejor calificado es de $k=8$, seguido de $k=2$. Pero se seleccionó un agrupamiento de 8 grupos para maximizar la representatividad del comportamiento del rendimiento de los lotes.

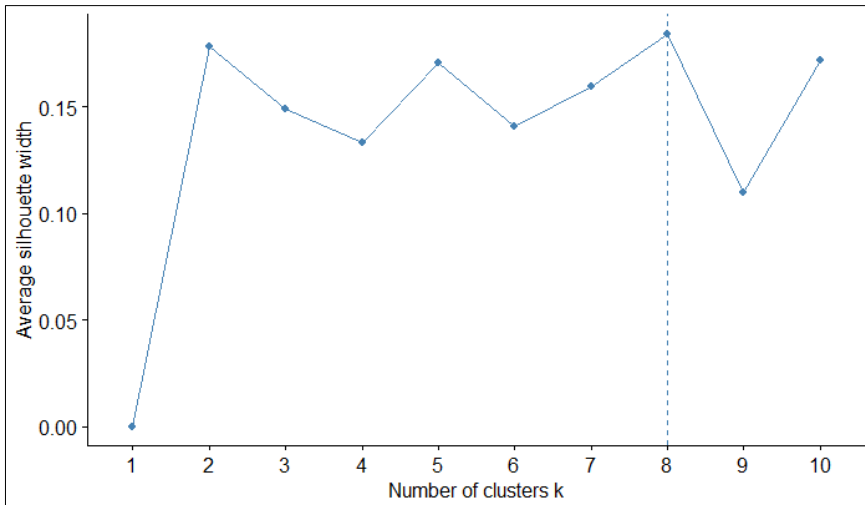


Figura 4.2 Determinación del número óptimo de clústeres mediante el método de Silhouette

Definido el número de clúster ($k=8$), se realizó el agrupamiento según se muestra en la figura 4 y de cada uno de los grupos se extrajo de forma aleatoria un lote. Preliminarmente los lotes seleccionados fueron 72, 77, 115, 137, 138, 144, 145 y 148, pero finalmente se sustituyó el 72 por el 118. Los motivos por lo que se seleccionó el 118, es que, es el lote que mejor representa el comportamiento promedio de la población analiza (figura 5) y además se excluyéndose el lote 72, debido a que presenta datos faltantes y a una caída anómala de su rendimiento.

Según se observa en la figura 4.4, el lote 144 representa la curva con mejor rendimiento de producción, en tanto el lote 77 es uno de los lotes que menor rendimiento productivo presentan. Aspecto, que es de interés en la selección de los lotes, porque, da algún grado de seguridad, de que el comportamiento del rendimiento de la plantación de palma de aceite es bien representado por el grupo de los 8 lotes seleccionados, debido a que abarca todo el espectro de la población en estudio. La ubicación de la lista final de los lotes que serán utilizados en la simulación con el modelo APSIM se muestra en la figura 4.3.

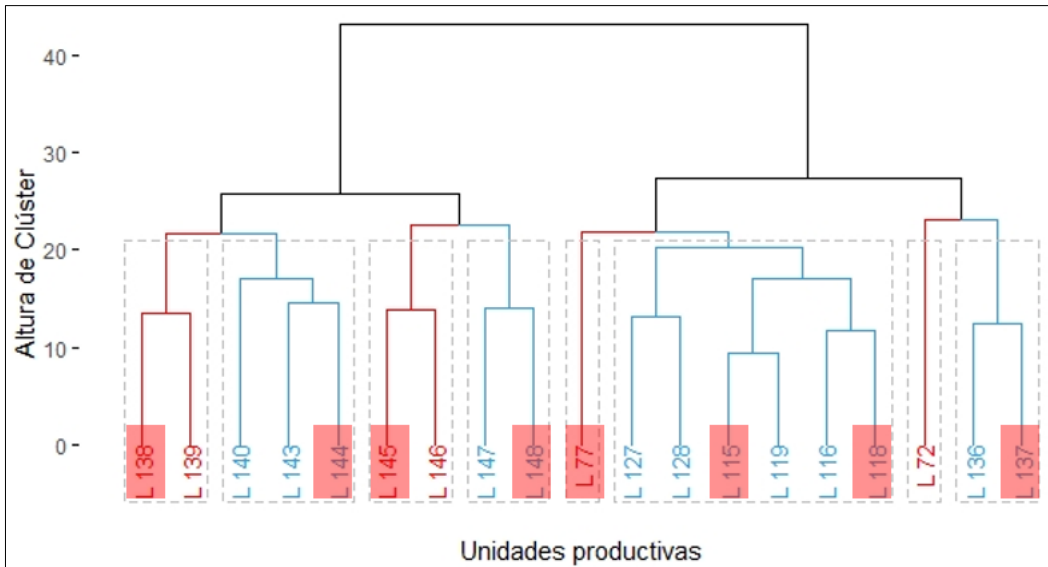


Figura 4.3 Clúster de los lotes de la plantación de palma de aceite según el rendimiento mensual. Los lotes seleccionados 77, 115, 118, 137, 138, 144, 145 y 148 se resaltan en el gráfico en color rojo.

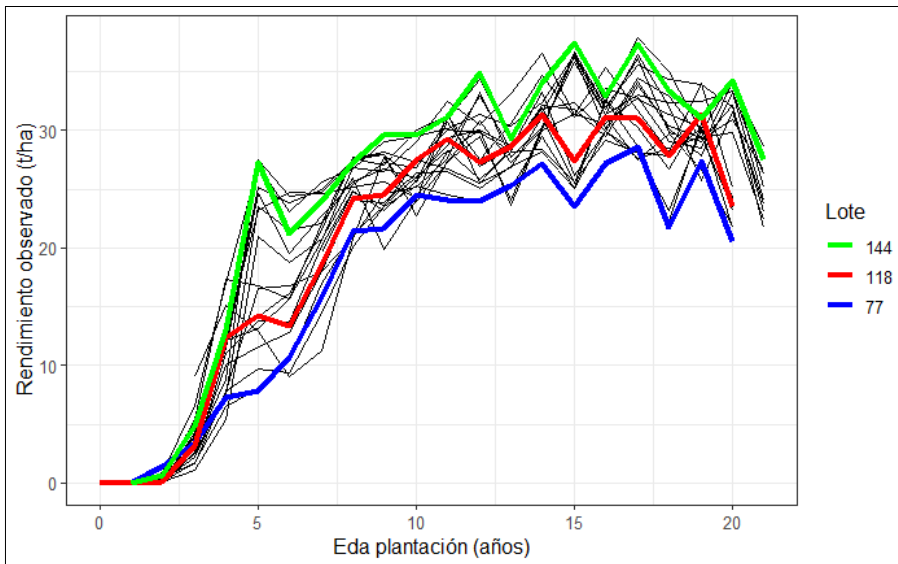


Figura 4.4 Comportamiento del rendimiento según la edad de la planta para los lotes cuya fecha de siembra esta entre 1995 y 1996.

4.2. Parámetros físicos del suelo

4.2.1. Textura del suelo

El rango de texturas de los lotes esta entre franco y arcillosa, siendo la textura más predominante la franco arcillo arenosa. Los rangos de densidad aparente se están entre

1.2 y 1.66 g/cm³, el promedio 1.28 g/cm³ con una desviación de 0.12 g/cm³ (Tabla 4.1 y Figura 4.5)

Tabla 4.1 Textura y densidad aparente de suelo de los lotes

Lote	Profundidad inicial (cm)	BD (g/cm ³)	Textura
77	0	1,59	Franco arcillosa
	25	1,35	Franco arcillo arenosa
	50	1,22	Franco arcillo arenosa
	90	1,22	Franco arcillo arenosa
115	0	1,20	Franco arcillo arenosa
	30	1,20	Franco arcillo arenosa
	60	1,30	Franco arcillo arenosa
	70	1,30	Franco arcillo arenosa
118	0	1,32	Franco arcillo arenosa
	30	1,33	Franco arcillo arenosa
	60	1,30	Franco arcillo arenosa
	60	1,35	Franco arcillo arenosa
	60	1,40	Franco arcillo arenosa
137	0	1,66	Franco arcillosa
	20	1,5	Franco arcilloso arenosa
	35	1,3	Franco arcillosa
	50	1,3	Franco arcillosa
	80	1,4	Franco arcillosa
138	0	1,03	Franco arcillosa
	20	1,03	Franco arcilloso arenosa
	35	1,20	Franco arcillosa
	50	1,20	Franco arcillosa
	80	1,20	Franco arcillosa
144	0	1,25	Franco arcillosa
	30	1,22	Franco arcillo arenosa
	50	1,22	Franco arcillo arenosa
	87,5	1,22	Franco arcillo arenosa
145	0	1,30	Franco arcillo arenosa
	30	1,22	Franco
	50	1,22	Franco
	87,5	1,22	Franco
148	0	1,34	Arcillosa
	30	1,15	Arcillosa
	50	1,20	Arcillosa
	87,5	1,30	Arcillosa

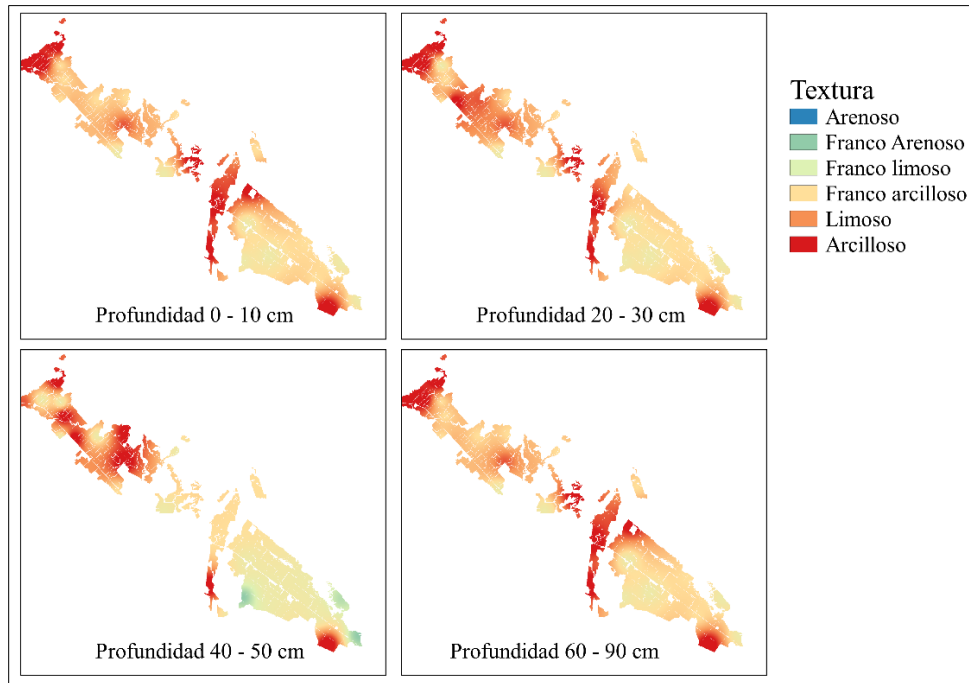


Figura 4.5 Modelo de distribución de la textura del suelo obtenido por el método IDW

4.2.2. Curvas de humedad de suelo

Los resultados completos obtenidos para cada lote sobre humedad del suelo se muestran en la figura 4.6. En términos generales, los lotes 138, 144, 145 y 148 son los que presentan un mayor rango de humedad disponible para la planta, en promedio $LL_{15}=0.12\text{mm/mm}$ y un $DULL=0.41\text{mm/mm}$. Los otros cuatro lotes restantes (77, 115, 118, 137) no muestran comportamientos similares entre sí, pero tanto el lote 77 y mucho más notorio el 137 muestran una drástica disminución en la humedad disponible en los primeros 25 cm de perfil de suelo.

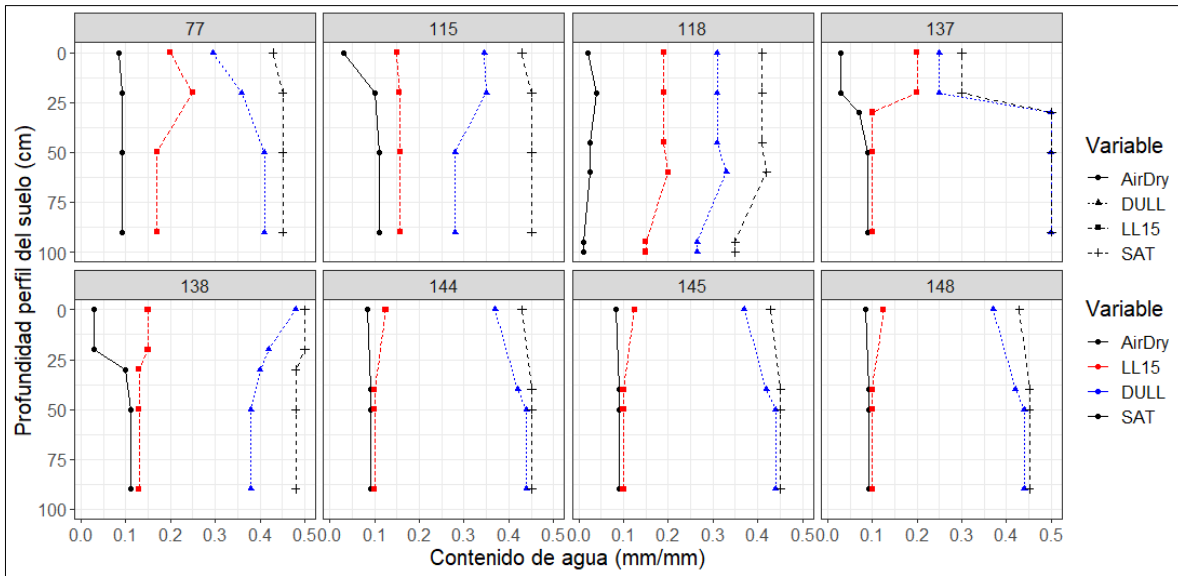


Figura 4.6 Curvas de humedad de suelo

4.3. Análisis de las series de tiempo de datos climatológicos de precipitación, temperatura y radiación

De las 14 estaciones pluviométricas con las cuales se contó en un inicio, se excluyeron 5 de ellas debido a que los datos faltantes superan el 10% de los totales. De las 9 estaciones que superaron este filtro, 7 tienen la serie de datos completa y 2 tienen menos del 0.5% de faltantes (Figura 4.7).

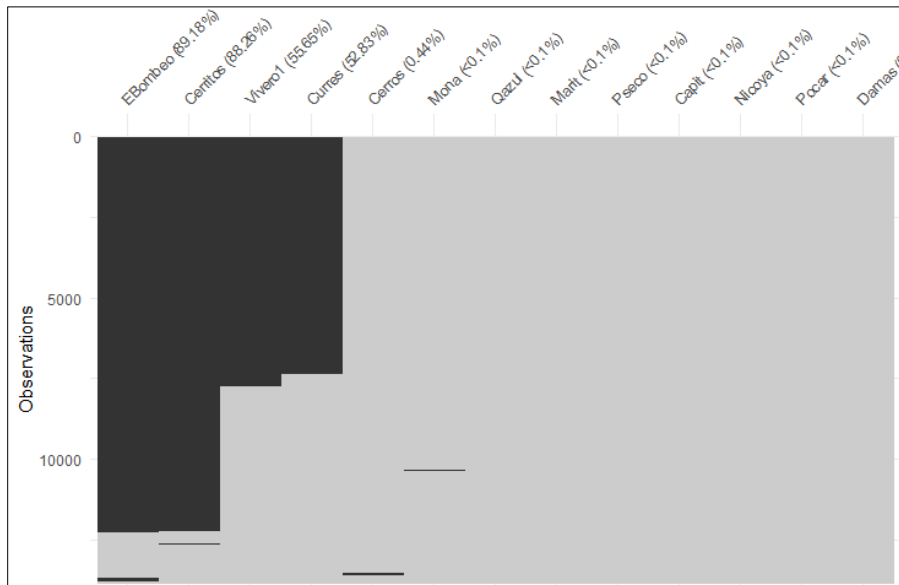


Figura 4.7: Visualización de datos faltantes en las estaciones pluviométricas.

Las estaciones que requieren de la aplicación de técnica para rellenar datos faltantes son Mona con 6 días sin información (0.04%) y Cerros con 61 días de datos faltantes (0.45%). Para lo cual se generaron ecuaciones de relleno a partir de las otras estaciones que mayor

coeficiente de correlación tuvieran (Figura xx). En el caso de Mona, las estaciones Marit, Capit generan coeficientes de Pearson superiores o iguales a 0,86. Para la estación Cerros, los coeficientes de correlación más altos son de 0,8 para las estaciones Qazul y Pocar (Figura 4.8).

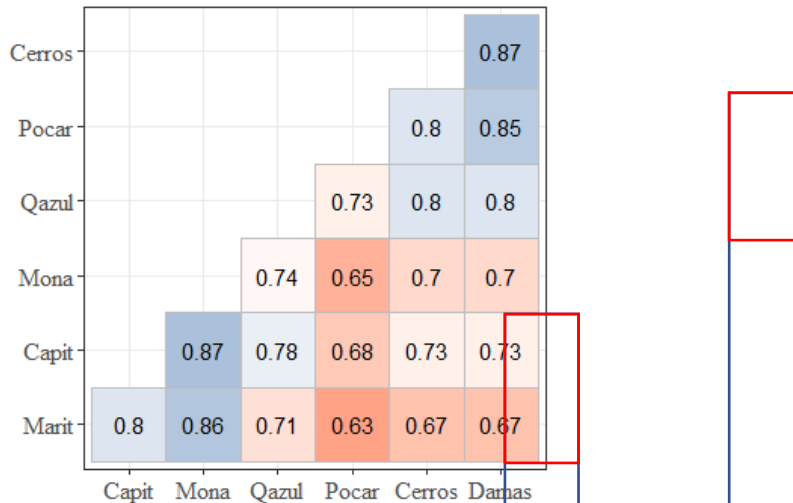


Figura 4.8 Proceso de generación de ecuaciones para relleno de datos faltantes de las estaciones pluviométricas Mona y Cerros, en función de las estaciones con las cuales guardan mayor correlación

El modelo generado para la Estación Mona obtuvo un coeficiente de determinación de 0.84 y un p-value de 2,2e-16 (ecuación 19), en tanto para la estación Cerros se obtuvo un r^2 de 0,74 y un p-value de 2,2e-16 (ecuación 20).

$$P_{Mona} = 0.44P_{Marit} + 0.48P_{Capit} + 0.39 \dots (19)$$

$$P_{Cerros} = 0,47P_{Qazul} + 0,53P_{Pocar} + 1,29 \dots (20)$$

Donde,

P = Precipitación diaria de la estación según se indica en el subíndice.

Una vez, se completaron de rellenar los datos faltantes de las estaciones, se generaron series de precipitación por interpolación (IDW) hacia los centroides de cada uno de los lotes seleccionados. Las estaciones utilizadas para la interpolación fueron Cerros, VDamas, Capit, Mona y Marit. Las estaciones y distancias que se utilizaron en cada lote se detallan en el Tabla 4.2.

Tabla 4.2 Distancias del centroide de los lotes en seleccionados hasta las estaciones de precipitación más cercanas

Lote	Distancia de la estación al centroide del lote (m)				
	Cerros	VDamas	Capit	Mona	Marit
77	1607,4				
115		3162,5	895,6		
118		2439,1	1624,7		
137		888,3		806,9	
138		1577,3		1321,0	
144				1254,2	2447,5
145				1296,3	2402,3
148				3071,6	

El promedio de precipitación acumulada anual de los lotes seleccionados esta entre 4537 mm a 3999 mm para el periodo de 1995 al 2015. El lote que en promedio mayor cantidad de lluvia acumulada recibe es el 148 con 4537 mm y el que menos cantidad recibe es el lote 77 con 3999mm. Las desviaciones estándar rondan entre 809 mm a 1069 mm (Figura 4.9).

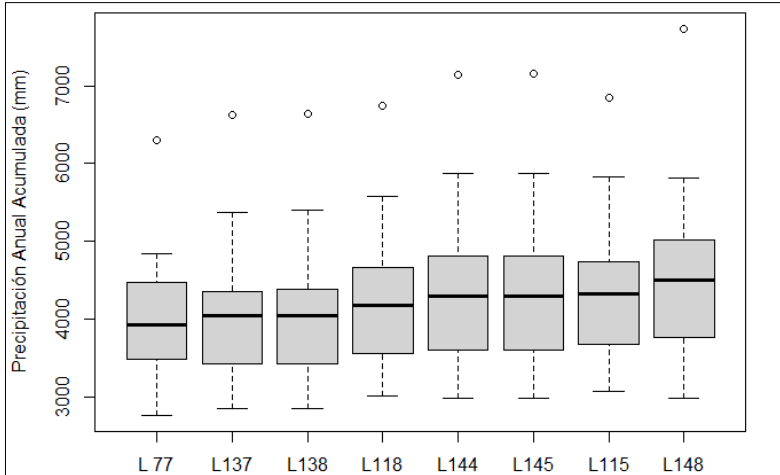


Figura 4.9 Distribución precipitación acumulada anual.

La distribución interanual del régimen de lluvias es la misma en todos los lotes, se caracteriza por tener tres meses secos enero, febrero y marzo, dos meses transitorios, diciembre de periodo lluvioso a seco y abril de periodo seco a lluvioso. En tanto, los meses con mayor precipitación van de mayo a noviembre, destacándose octubre como el de mayor precipitación promedio (Figura 4.10).

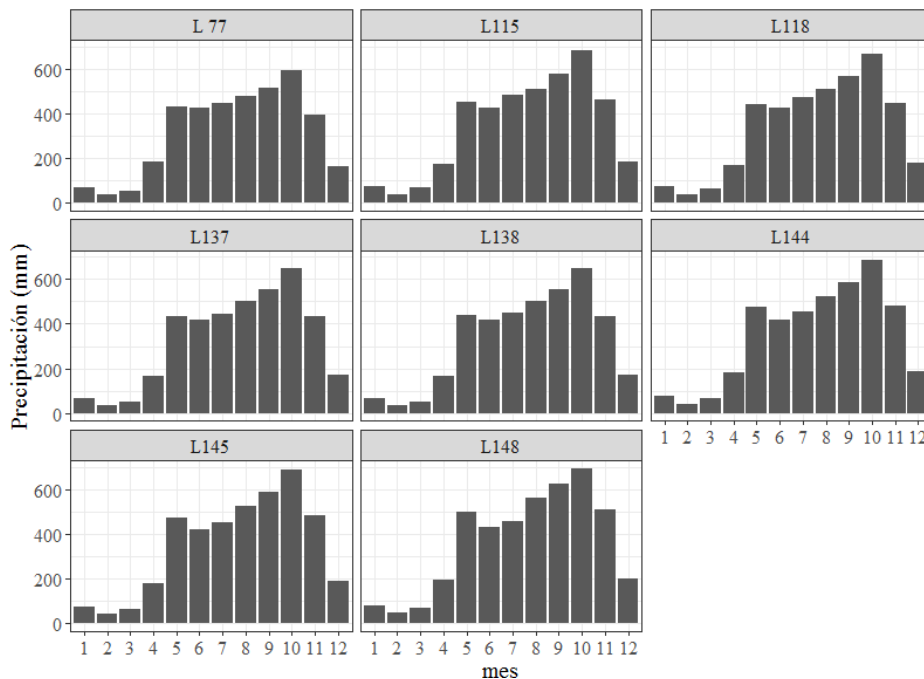


Figura 4.10 Régimen de lluvias en para cada lote en estudio.

Los datos de temperatura y radiación no fueron interpolados a cada lote por contarse de solamente dos estaciones, por lo que se utiliza una misma serie de datos para todas las unidades productivas. Se determinó que el mes en donde se presentan las temperaturas máximas es marzo con un promedio de 32.47°C y las mínimas en enero con un promedio de 21.90°C, la media anual ronda 26.92°C (Figura 4.11). La máxima observada se presentó en febrero de 1995 alcanzando 36.2°C y la mínima en noviembre del 2013 con 16.30 °C.

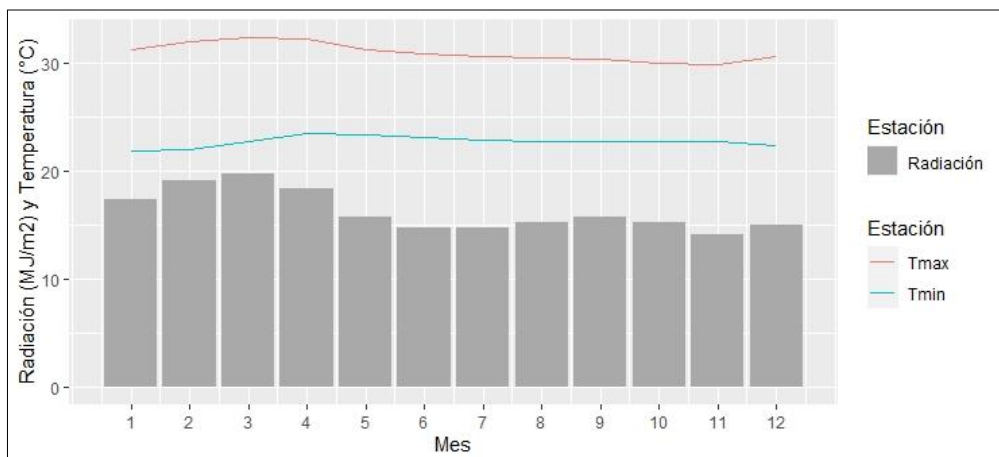


Figura 4.11 Régimen mensual de temperatura y radiación.

En cuanto a la radiación, de enero a abril son los meses que mayores valores de radiación se muestran, entre los cuales, marzo es el mes de mayor incidencia alcanzando valores de 19.80 MJ/m²/d/d. De mayo a diciembre se presentan valores mínimos con un comportamiento homogéneo (Figura 4.11).

La base de datos climática, resultante en cada lote contiene información de precipitación, que como se indicó anteriormente es interpolada a partir de las ubicaciones y datos de cada una de las estaciones pluviométricas utilizadas. En tanto las temperaturas máximas, mínimas y radiación es una misma serie de datos.

4.4. Resultados del modelo APSIM en cada uno de los lotes.

Para cada uno de los lotes seleccionados 77, 115, 118, 137, 138, 144, 145 y 148, se crearon modelos en la herramienta APSIM, a partir de la información de suelo y clima mostrados en los puntos anteriores.

Los modelos generaron coeficientes de evaluación de NSE entre 0.64 y 0.74, coeficientes de determinación entre 0.65 y 0.75, con p-value menores de 1.3e-5. Los coeficientes de error en promedio son para RMSE de 5.86 t/ha y un MAE de 4.90 t/ha (Figura 4.12).

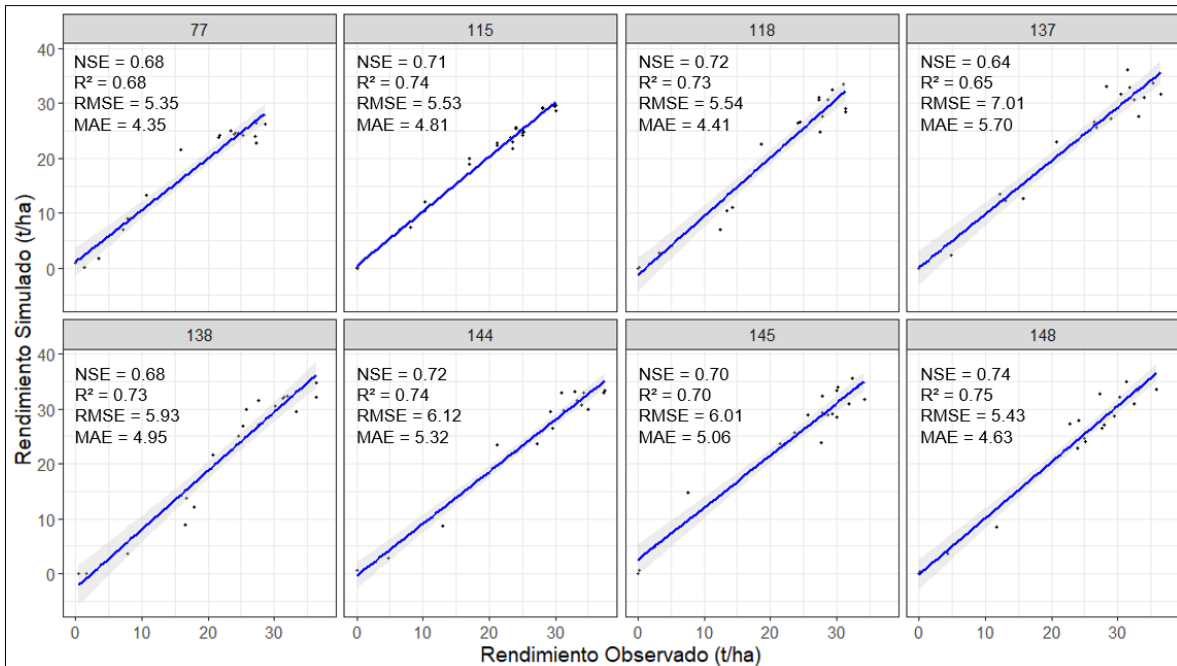


Figura 4.12 Rendimiento observado comparado contra el simulado de los modelos APSIM para cada lote seleccionado en el periodo 1996-2015

El modelo APSIM en el lote 137 fue donde obtuvo los mayores coeficientes de error (RMSE=7.01 t/ha, MAE=5.70 t/ha) y los coeficientes NSE y R² más bajos. En tanto, en el lote 148 la simulación mostro los mejores resultados según los coeficientes tanto de error como de eficiencia (NSE=0.74, R²=0.75, RMSE=5.43 t/h, MAE=4.63 t/ha).

4.5. Escenario de cambio climático

El escenario de cambio climático utilizado fue realizado por Hugo Hidalgo para el período de 1974 al 2099. se observó que existen diferencias para el periodo observado de 2000 al 2015 entre el escenario de cambio climático y los datos de precipitación y temperatura, por lo que se realizó una corrección, tanto para los datos de temperatura como para los datos de precipitación Según se muestra a continuación.

4.5.1. Ajustes ECC según periodo observado

La media de la temperatura máxima del escenario cambio climático (maxT_ECC) para el período 2000 al 2015 es de 33.44°C (Tabla 4.3), en tanto, la media de los datos observados de temperatura máxima (maxT_obs) para ese mismo período es de 31.07°C, lo que genera una separación de 2.37°C y es apreciable en la figura xx. También existen diferencias en la desviación estándar que para maxT_ECC es de 1.7°C, en tanto para maxT_obs es de 1.37°C. Estas diferencias, pero en menor magnitud, también ocurren al comparar las temperaturas mínimas. Por lo que, para realizar este ajuste de modo que se igualen tanto las medias como las desviaciones estándar, se realizó el procedimiento descrito en la metodología en el punto 3.4, cuyos resultados se muestran en la figura 4.13.

Tabla 4.3 Diferencia en la media y desviación estándar entre el ECC y datos observados para el periodo 2000 - 2015

Temperatura		Mean	SD
Máxima	Observado	31,07	1,37
	ECC	33,44	1,70
Mínima	Observado	22,93	1,08
	ECC	22,36	1,42

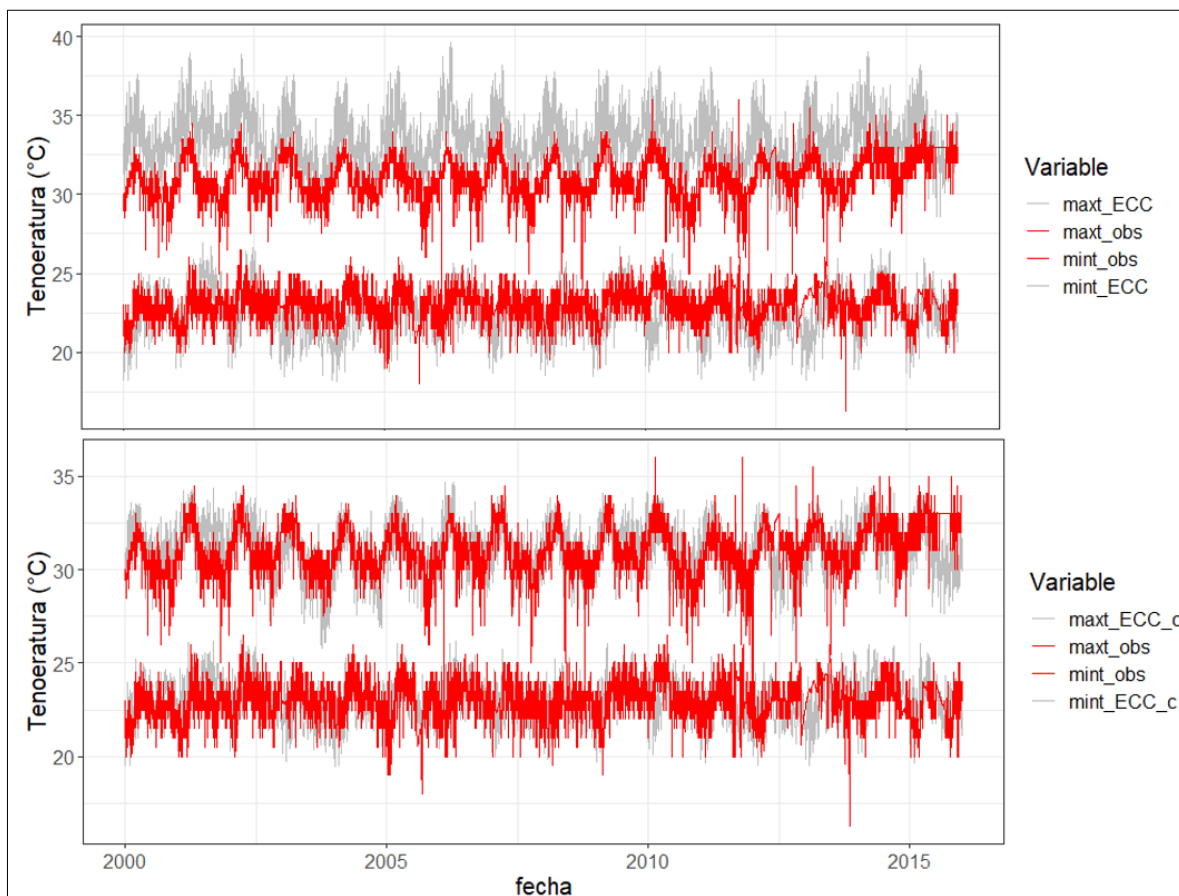


Figura 4.13 Resultados del ajuste del escenario de cambio climático en función del periodo observado

Los datos corregidos del ECC (maxT_ECC_c y minT_ECC_c) se obtuvieron hasta el año 2099. Similar procedimiento se realizó con los datos de precipitación, pero considerando el comportamiento del regimiento de precipitaciones para realizar el ajuste. Al obtener las diferencias entre el ECC base de precipitación (ECC_Base) y el periodo observado (2000-

2015) de forma mensual (Figura 4.14), se determinó que ECC_Base presentaba promedios mensuales de precipitación inferiores que los observados.

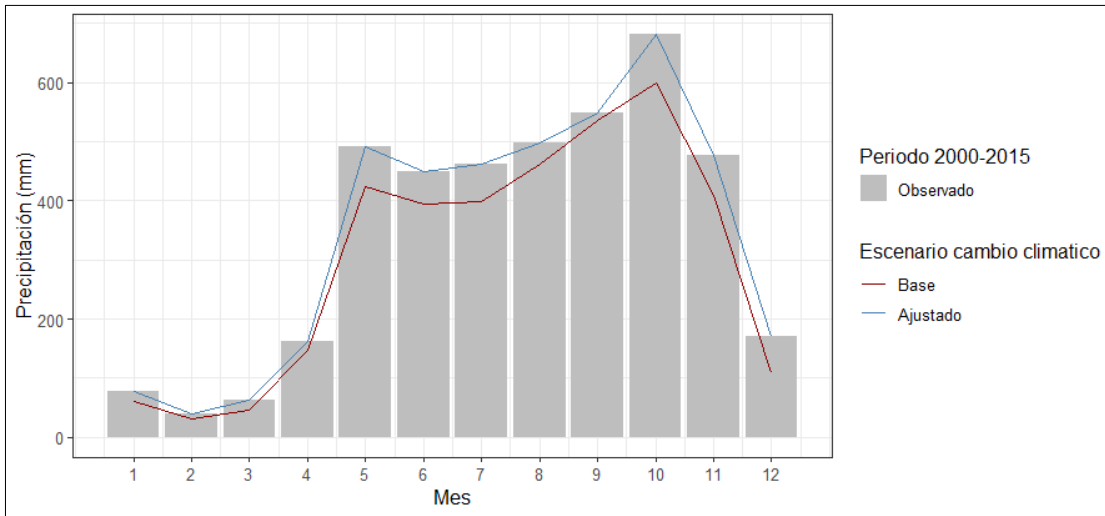


Figura 4.14 Ajuste del ECC de precipitación

Una vez realizados las correcciones de sobre el escenario de cambio climático, fue necesario generar de forma indirecta datos de radiación a partir de variables conocidas (maxT, minT y precipitación). Debido a que estas variables climáticas son entradas requeridas en el modelo APSIM. Para tal fin crearon modelos de machine learning cuyos resultados se describe a continuación.

4.5.2. Generación de serie de radiación

A partir de las series de datos climáticos observados 2000 – 20015 (mixto, minT, precipitación y radiación) y el número de día del año, se crearon dos tipos de modelos de machine learning un Random Fores y un Neural Network. De tal forma que el mejor de ellos pueda ser utilizado para generar la serie de datos de radiación del ECC.

En la construcción de los modelos se estandarizaron las variables utilizadas (figura 4.15), debido a que el rango de valores no es uniforme, sobre todo por las variables, precipitación (rain_obs) y número de día del año (dy).

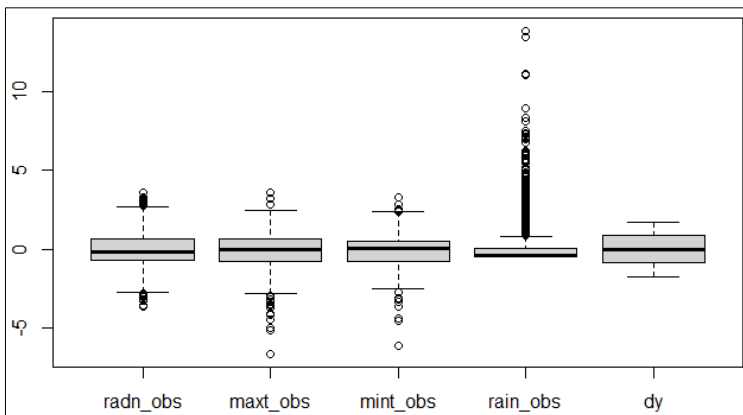


Figura 4.15 Comportamiento del rango de valores de las variables estandarizadas

En la construcción de los modelos se contempló un 70% de los datos para el entrenamiento y un 30% para la validación. La mejor calibración de parámetros para RF es un ntree de 1000 y un mtry de 4. En tanto, para el NN se utilizó un threshold de 0.0001, stepmax de 1×10^6 y el algoritmo utilizado fue el rprop+. Según los resultados de los modelos se determinó que tanto en el proceso de calibración como en la validación el RF fue el que mejores valores de eficiencia y error obtuvo respecto al NN (Figura 4.16).

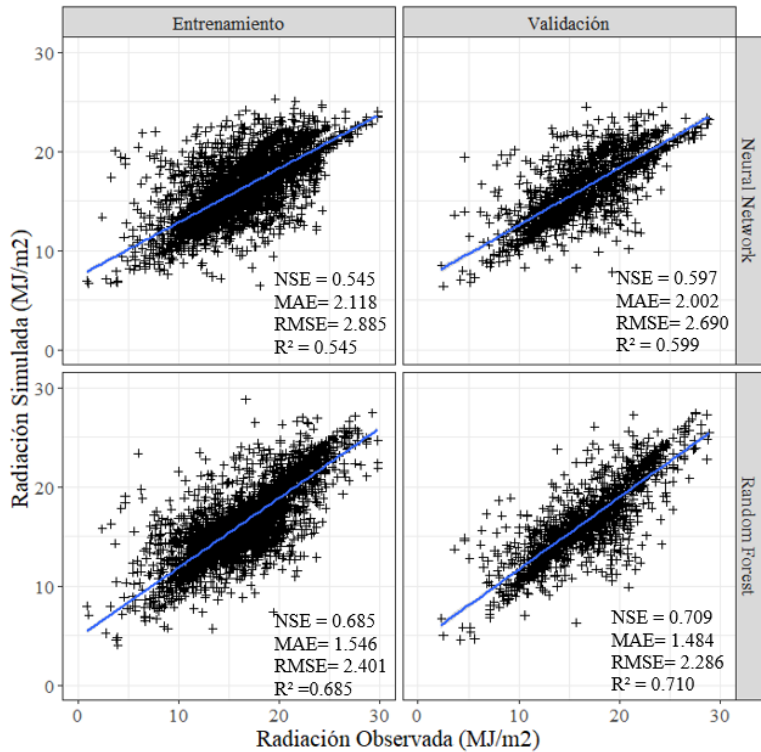


Figura 4.16: Resultados de los modelos para cálculo de radiación

En la validación RF tuvo un rendimiento bueno con un NSE = 0.709 y un R² = 0.710, con errores de MAE=1.484 MJ/m²/d/d y un RMSE = 2.286 MJ/m²/d/d (Figura xx). Con el modelo RF y utilizando las variables del ECC 2000 a 2099 se generó para ese mismo periodo los datos de radiación. Esta serie de datos en adelante se divide en tres periodos; línea base (2000-2019), ECC1 (2040-2059) y ECC2 (2080-2099) (Figura 4.17).

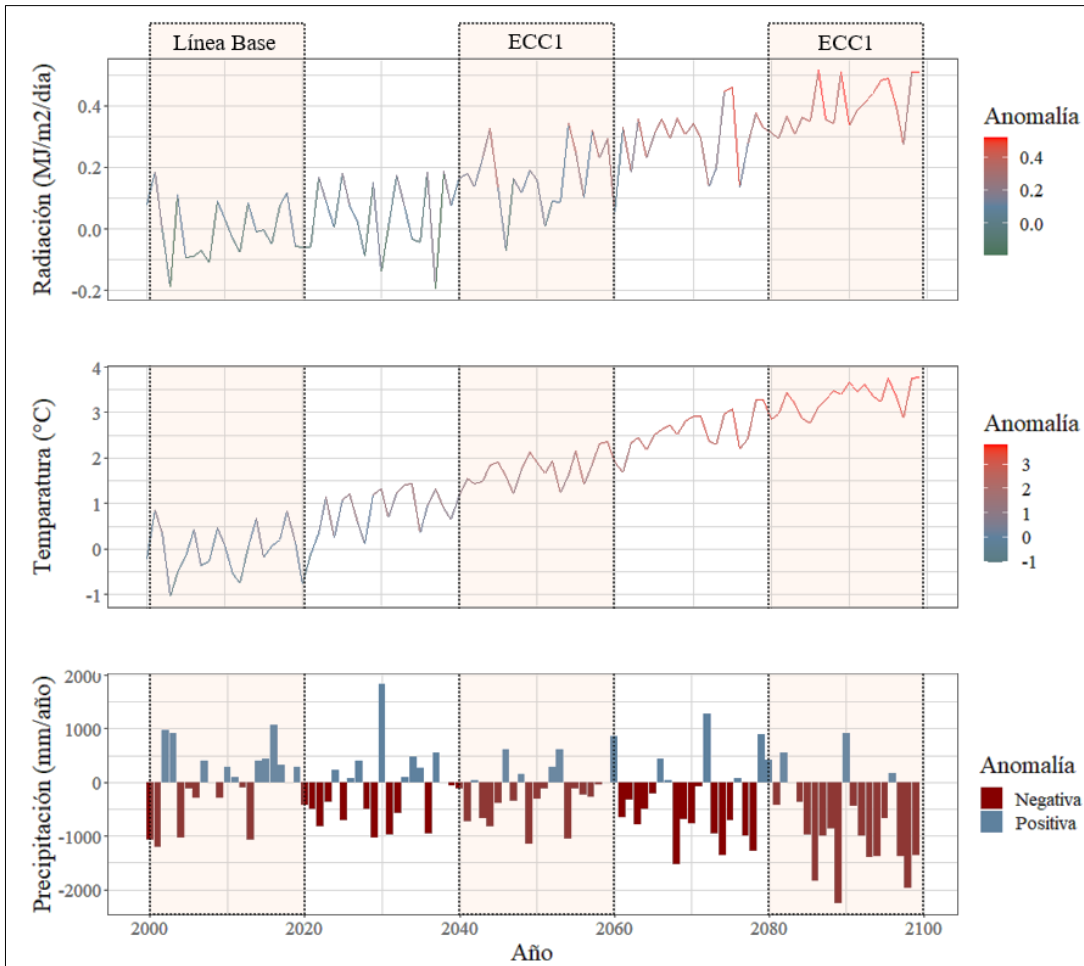


Figura 4.16 Anomalía de las variables precipitación, temperatura y Radiación de los ECC respecto a la línea base.

La línea base es el periodo contra el cual se compararán los escenarios de cambio climático. En este periodo se tienen una precipitación anual acumulada de 4224.02mm, la temperatura promedio es de 27.07°C y la radiación promedio es de 16.20 MJ/m²/d/d (Tabla 4.4).

Tabla 4.4 Comparación de las variables climatológicas para cada escenario

Escenario	Periodo	Precipitación	Temperatura			Radiación
			máxima	mínima	Promedio	
Base	2000-2019	4224,02	31,15	23,00	27,07	16,20
ECC	2040-2059	3989,67	32,87	24,74	28,81	16,38
	2080-2099	3461,33	34,46	26,32	30,39	16,60

El ECC1 presenta una disminución en la precipitación acumulada anual de 234.35 mm, esto es equivalente a un 5.55% de reducción. La temperatura promedio aumentara en 1.73°C y al igual que la radiación con un incremento de 0.18 Mj/ m² respecto a la línea Base. El CC2 tienen cambios más fuertes, con una disminución en la precipitación de 762.69 mm (-

18.06%), conjunto a un aumento de la temperatura de 3.31°C y de la radiación de 0.40 Mj/m².

4.5.3. Simulación de los efectos de los escenarios ECC1 y ECC2 sobre el rendimiento del cultivo de palma de aceite

A partir de los modelos APSIM calibrados para los lotes 77, 115, 118, 137, 138, 144, 145 y 148, se crearon tres conjuntos de simulaciones para Línea base (2000-2019), ECC1 (2040-2059) y ECC2 (2080-2099). Con las cuales se obtuvieron los rendimientos de estas áreas tendían como respuesta a las series de datos de maxT, minT, precipitación y radiación del escenario de cambio climático utilizado. Los resultados de estos conjuntos de simulaciones se compararon y se obtuvieron las diferencias entre los ECC1 y ECC2 respecto a la Línea Base.

En cada conjunto de simulaciones se realizaron las corridas de los 8 lotes seleccionados, generando un rango de respuesta esperada según el año (Figura 4.18). Para las simulaciones de la Línea base se obtuvo que en promedio el rendimiento fue de 21.63 t/ha/año calculadas para todo el periodo de siembra. Al comparar, se determinó que en ambos escenarios se dio una disminución de rendimiento promedio, para el ECC1 la disminución fue de 1.7 t/ha/año y para el ECC2 fue de 8.19 t/ha/año.

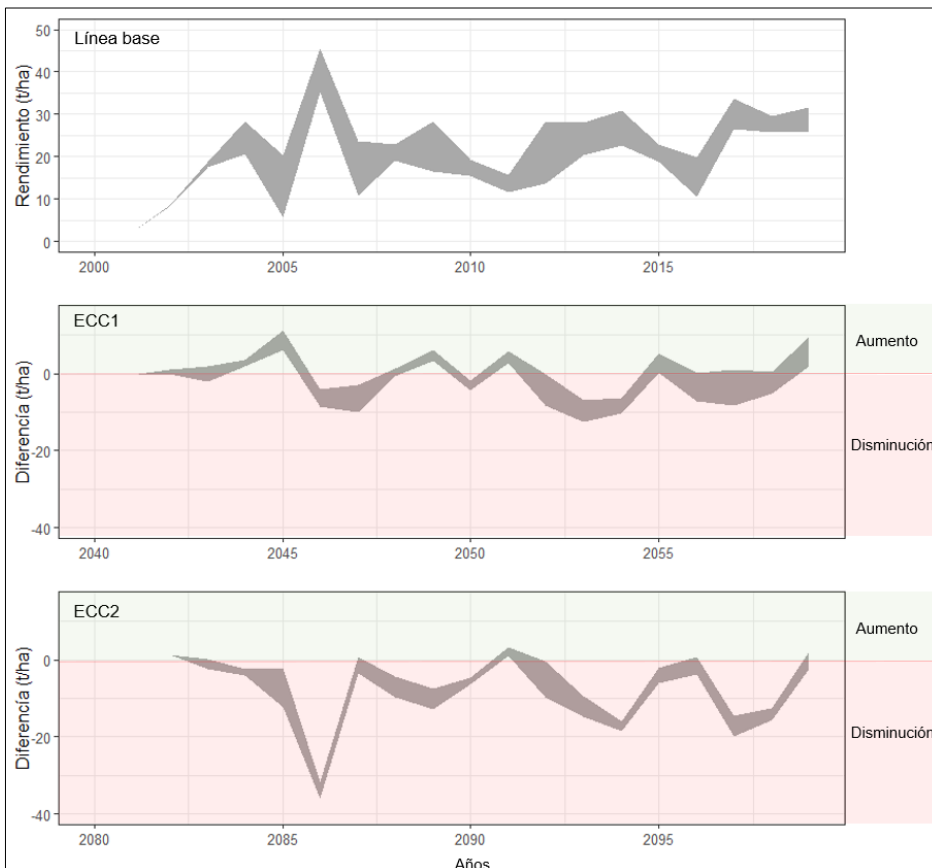


Figura 4.17 Efecto del cambio climático sobre el rendimiento de palma de aceite.

Al determinar que el rendimiento de la plantación de palma de aceite es afectado por el cambio climático, se volvieron a realizar las simulaciones, pero aplicando riego como

medida de adaptación, para lo cual, se contempló una profundidad de aplicación de 0.75cm, la eficiencia de aplicación del 80% y porcentaje de agotamiento variado, de tal forma que el rendimiento en los ECC se recupere y al menos iguale el observado en la Línea Base. La evaluación se realizó comparando los promedios mensuales del agua disponible en el suelo (ADS) con aplicación y sin aplicación de riego para cada lote y para cada ECC.

En el ECC1, en general, se observó que el agua disponible en el suelo (ADS) difiere mayoritariamente en los meses secos de enero, febrero, marzo y abril. En tanto en los meses de mayor precipitación de mayo a noviembre e incluyendo en algunos casos diciembre, como mes de transición, se determinó que el comportamiento de la humedad de suelo muy similar (Figura 4.19).

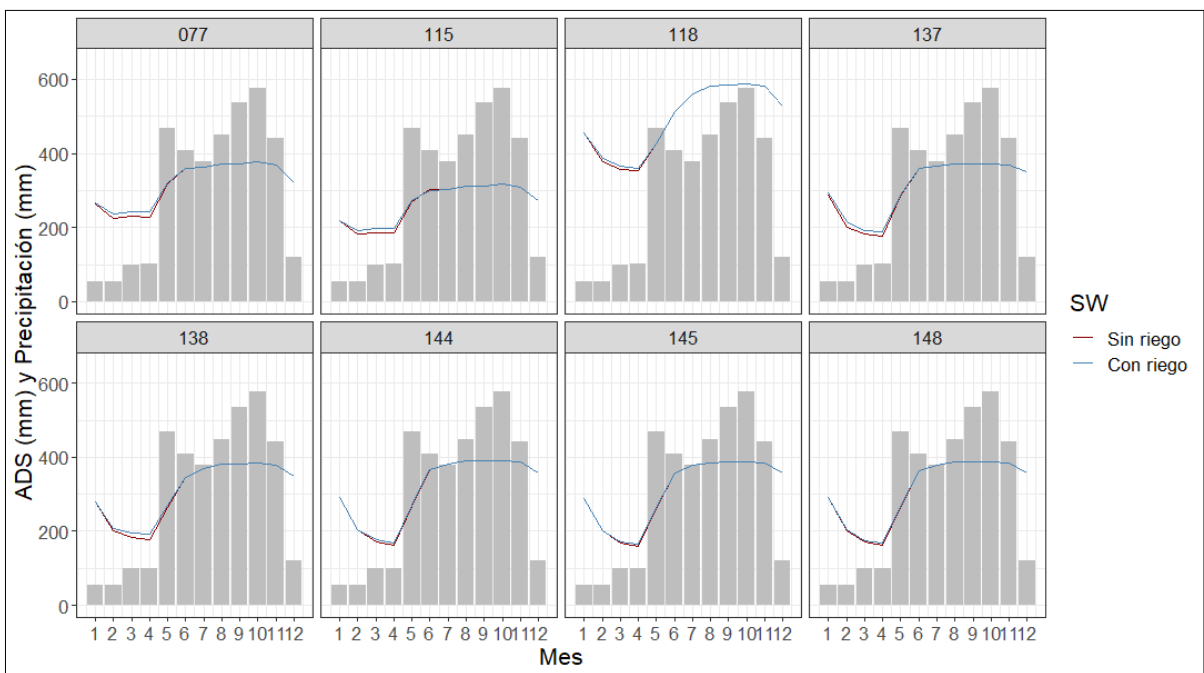


Figura 4.18 Variación sobre el agua disponible en el suelo producto de la aplicación o no de riego bajo el ECC2.

La diferencia entre las curvas de ASD con y sin riego (ΔADS) se puede explicar como el efecto del riego sobre la humedad del suelo necesario para mantener los niveles de rendimiento de la Línea base, la cual varían dependiendo del mes. El lote que mayores requerimientos hídricos presenta es el 137, con diferencias máximas de 13.82 mm en marzo y con aplicación de riego de diciembre a julio. El lote 145 es el que menos demanda hídrica presenta, requiere riego los meses febrero a mayo, con máximos en la ΔADS en abril de 5.5 mm en el promedio mensual.

En el ECC2, en general, se observó que el agua disponible en el suelo (ADS) con y sin riego difiere hasta en meses lluviosos inclusive hasta el mes de agosto y con diferencias mucho mayores comparadas con las obtenidas en el ECC1. Además, es muy notable la disminución de la precipitación promedio mensual, con especial atención al mes de junio

debido a que la disminución es en promedio de 140 mm, pero en donde la diferencia respecto al mes anterior y posterior se acentúa respecto al ECC1 (Figura 4.20).

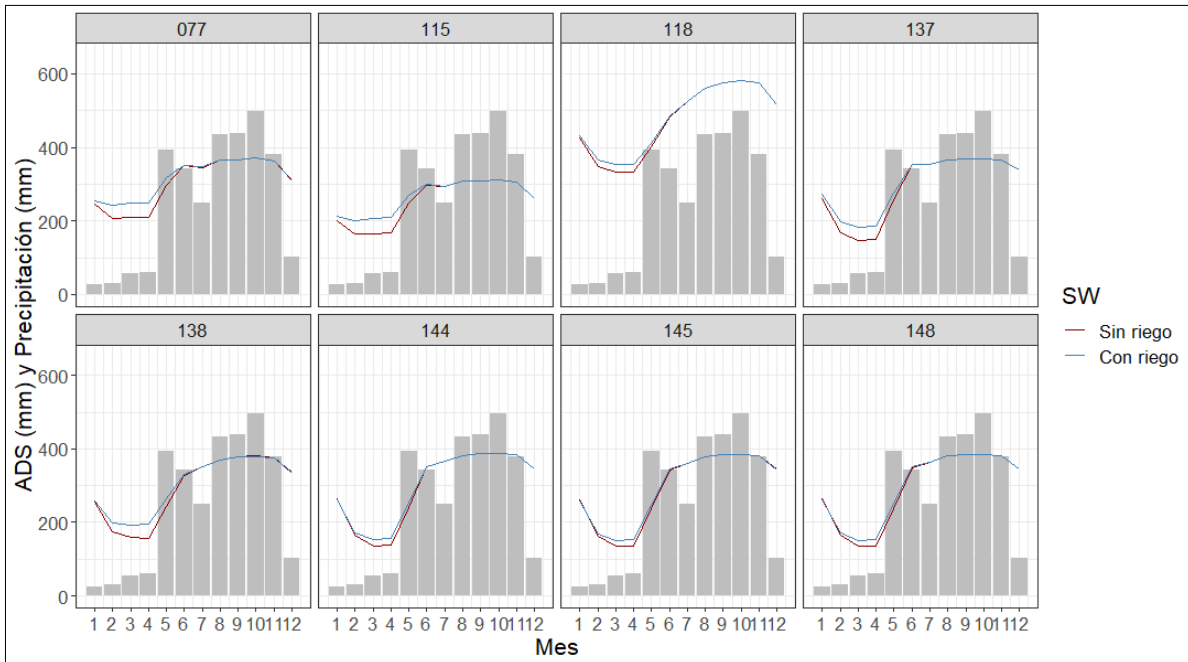


Figura 4.19 Variación sobre el agua disponible en el suelo producto de la aplicación o no de riego bajo el ECC2.

Para ECC2, los requerimientos de riego aumentan, con un ΔADS máximo para el mes de marzo de 42.38 mm en el lote 115, en tanto el lote 148 presenta el pico de menor magnitud con $\Delta ADS = 18.53$ mm. De forma general, el riego debe ser implementado de enero a junio y en algunos casos es requerido en julio, agosto o diciembre para poder mantener los rendimientos obtenidos en la Línea Base (Figura 4.20). En la Figura 4.21 se muestra el rango de ΔADS de los 8 lotes para los ECC1 y ECC2.

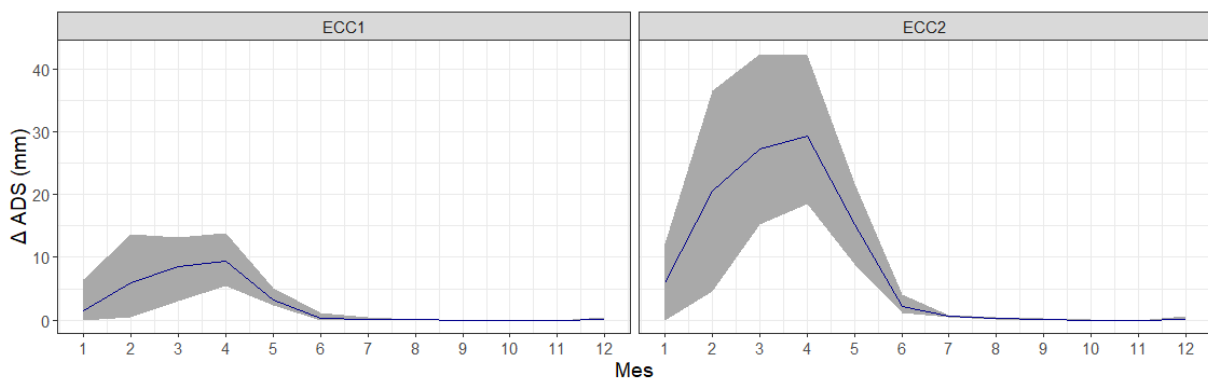


Figura 4.21: Rango de diferencias obtenidos en los 8 lotes de estudio entre el ADS bajo condiciones de cultivo sin riego y con riego para los ECC1 y ECC2.

5. DISCUSIÓN

Utilizando el método de Silhouette se determinó que ocho es el número óptimo de agrupamientos de un total de 19 lotes según la variable rendimiento de fruta. Este proceso de agrupación se complementó con una selección al azar de un lote representante para cada grupo. Los lotes seleccionados fueron 77, 115, 118, 137, 138, 144, 145 y 148, permitiendo obtener una muestra de unidades de producción que abarcara todo el espectro del comportamiento del rendimiento del cultivo de palma de aceite (Taylor, Dror, & Steinberg, 2012).

En los lotes seleccionados la textura predominante va de franco arcillo arenosa a Franco arcillosa, con excepción del lote 148 que es arcilloso y el lote 145 que presenta dentro de su perfil texturas francas. Los rangos de densidad aparente de suelo obtenidos en laboratorio en general concuerdan con los valores teóricos expresados por algunos autores entre 1.00 y 1.4 g/cm³ ((ORGANIZACIÓN DE LAS NACIONES UNIDAS PARA LA AGRICULTURA Y LA ALIMENTACIÓN, 2009). Sin embargo, los lotes 77 y 137 en las capas más superficiales (textura franco arcillosa), tienen valores de densidad aparente considerados como altos de 1.59 y 1.66 g/cm³ respectivamente (Rai et al., 2017). Esto podría indicar que existe algún factor como compactación, que disminuye la porosidad del suelo que afectando la densidad aparente y que genera la disminución en valores de capacidad de campo del estrato de suelo entre de 0 a 25 cm de profundidad, que son de 0.295 mm/mm para el lote 77 y 0.25 mm/mm para lote 137, relativamente más bajos que el resto del perfil (Salamanca & Siavosh, 2005).

Valores promedios de temperatura máxima (31.07°C), temperatura mínima (22.93°C) y precipitación (4537 mm a 3999 mm) utilizados para realizar las correcciones y ajustes en los escenarios de cambio climático y posteriormente para la calibración de los modelos APSIM concuerdan con los indicados por el Instituto Meteorológico de Costa Rica y otros autores (Wright, 2009, Comité Regional de Recursos Hidráulicos [CRRH], 2008).

De igual forma, sucede con la radiación promedio de 15.52 MJ/m²/d que está entre los rangos indicados por Wright (2009). Sobre esta misma variable, los resultados del modelo RF utilizado para generar los datos de radiación del ECC obtuvo coeficientes de error y de evaluación de RMSE = 2.286 MJ/m²/d, MAE = 1.484 MJ/m²/d, NSE = 0.709 y R² = 0.710 que son calificados como buenos (Moriassi et al., 2007) y comparables con modelos generados por otros autores como Feng et al (2019), quien a partir de datos de temperaturas máximas y mínimas generó modelos con técnicas de machine learning con valores de NSE entre 0.707 a 0.878 y RMSE de 0.188 a 3.939 MJ/m²/d.

El hecho que se logró realizar una calibración del modelo APSIM en la totalidad de los 8 lotes seleccionados con coeficientes de eficiencia y error aceptables y coherentes con los resultados de Huth et al (2014), permite asegurar que existe fiabilidad sobre las variables de entrada utilizadas (precipitación, radiación, maxT, minT, BD, AirDry, LL15, DULL y SAT). En promedio los coeficientes de eficiencia y error obtenidos por los modelos fueron de NSE de 0.699, R² de 0.716, RMSE de 5.865 t/ha y MAE de 4.903 t/ha en tanto los obtenidos por los desarrolladores del módulo de palma de aceite en APSIM son de NSE de 0.560, R² de 0.620 y un MAE de 3.490 t/ha (Huth et al, 2014).

Los escenarios de cambio climático de Hidalgo et al (2015) ajustados para la zona de estudio a partir de datos observados de maxT, minT y precipitación. Estos datos muestran que a mediados del siglo XXI (ECC1) se tendrá una disminución de las lluvias acumuladas anuales del 5.55% y un aumento en la temperatura promedio de 1.73°C, en tanto, para finales de siglo (ECC2) el escenario sería más drástico con una disminución de la precipitación del 18.06% y un aumento de la temperatura de 3.31°C, diferencias calculadas respecto al periodo 2000-2019. Este escenario aun cuando está basado en un estudio del 2015, al compararse con trabajos más recientes como el de Alvarado Gamboa (2021) se muestran similitudes en el déficit en la precipitación, en el aumento de temperatura y en cómo estos se vuelven más drásticos conforme avanza el siglo, importante acotar que también en ambos estudios se determina acentuación de la canícula para el periodo del ECC2.

Utilizando los modelos APSIM calibrados para cada uno de los 8 lotes seleccionados pero variando los datos de clima y periodos de siembra para adaptarlos a Línea Base (2000-2019), ECC1 (2040-2059) y ECC2 (2080-2099), se logró determinar que las variaciones climáticas menos favorables generaran una disminución en el rendimiento de la producción del cultivo de palma de aceite. De tal forma, que para el periodo ECC1 la reducción de la producción es del 7.86% y para el ECC2 del 37.86% respecto de la producción media de la Línea Base (21.63 t/ha/año). En comparación, en Malaysia se determinó que aumentos de temperatura entre 1 y 4°C producirán disminuciones en la producción entre un 10% y 41% (Sarkar, Begum, & Pereira, 2020), lo cual, muestra similitudes con los porcentajes de disminución de rendimiento obtenidos en nuestro estudio.

Como medida de adaptación se cuantificó el Δ ADS, requerido en el ECC1 y ECC2 para mantener los rendimientos obtenidos en la Línea base. Esto se realizó simulando los escenarios con el software APSIM y el módulo de irrigación. Para el ECC1 se obtuvo que en promedio para los 8 lotes se requiere riego los meses de enero, febrero, marzo, abril y mayo con aplicaciones que generen un Δ ADS de 1.60, 5.98, 8.67, 9.60 y 3.34 mm respectivamente. En el ECC2 los requerimientos hídricos son mayores demandando riego los meses de enero a junio con láminas que en promedio alcancen un Δ ADS de 16.76 mm y máximos de 29.27 mm. En ambos casos, en el análisis interanual existe un incremento en los requerimientos hídricos del cultivo de palma de aceite conforme se establece la época seca. Además se determinó, que conforme el ECC se vuelva más intenso, con disminución de precipitación y aumento de temperatura, el agua disponible en el suelo disminuye aumentando los requerimientos de riego del cultivo de palma de aceite (Woznicki & Nejadhashemi, 2015).

6. CONCLUSIONES Y RECOMENDACIONES

En este estudio de adaptabilidad al cambio climático del cultivo de palma mediante la aplicación de riego se obtuvieron los siguientes resultados:

Del área de estudio analizada, 19 lotes de palma de aceite cumplieron con los requisitos año de siembra (1995 o 1996) y datos faltantes propuestos (>5%), determinando que 8 es el número óptimo de agrupamientos, según la homogeneidad de series de datos de rendimiento de cada uno de ellos al utilizar el método de Silhouette. Finalmente, de forma aleatorio se seleccionaron los lotes 77, 115, 118, 137, 138, 144, 145 y 148, quedando un representante por cada grupo.

Sobre estos 8 lotes se generaron de forma exitosa modelos de simulación de producción agrícola (APSIM), con coeficientes de eficiencia y error aceptables y coherentes con los resultados de Huth et al (2014), los cuales permite asegurar la fiabilidad sobre las variables de entrada utilizadas (precipitación, radiación, maxT, minT, BD, AirDry, LL15, DULL y SAT). En promedio los coeficientes de eficiencia y error obtenidos por los modelos fueron de NSE de 0.699, R² de 0.716, RMSE de 5.865 t/ha y MAE de 4.903 t/ha en tanto los obtenidos por los desarrolladores del módulo de palma de aceite en APSIM son de NSE de 0.560, R² de 0.620 y un MAE de 3.490 t/ha.

Se estimó que a mediados del siglo XXI (ECC1) se tendrá una disminución de las lluvias acumuladas anuales del 5.55% y un aumento en la temperatura promedio de 1.73°C, en tanto, para finales de siglo (ECC2) el escenario sería más drástico con una disminución de la precipitación del 18.06% y un aumento de la temperatura de 3.31°C respecto a las condiciones del periodo 2000-2019. Estableciendo que el cambio climático producirá condiciones desfavorables para el cultivo de palma de aceite, disminuyendo en el rendimiento de la producción del cultivo que para periodo ECC1 es del 7.86% y para el ECC2 del 37.86% respecto de la producción media de la Línea Base (21.63 t/ha/año).

Finalmente, según las simulaciones realizadas se estableció que el riego permite adaptar las condiciones de agua disponible en el suelo, al punto que permita mantener los rendimientos de la Línea base del cultivo de palma para los escenarios de cambio climático propuestos (ECC1 y ECC2). En ambos casos, en el análisis interanual mostro un incremento en los requerimientos hídricos del cultivo de palma de aceite conforme se establece la época seca. Además, que conforme el ECC se vuelva más intenso el agua disponible en el suelo disminuye aumentando los requerimientos de riego del cultivo de palma de aceite.

7. BIBLIOGRAFÍA

- Akram, H., Levia, D. F., Herrick, J. E., Lydiasari, H., & Schütze, N. (2022). Water requirements for oil palm grown on marginal lands: A simulation approach. *Agricultural Water Management*, 260, 107292. <https://doi.org/10.1016/j.agwat.2021.107292>
- Alonso Ramírez Cover. (2007). Conflictos socioambientales y recursos hídricos en Guanacaste; una descripción desde el cambio en el estilo de desarrollo (1997-2006). *Anuario de Estudios Centroamericanos*, 33(1-2), 359-385. <https://doi.org/10.2307/40682777>
- Alvarado, A., & Escobar, R. (2017). Mejoramiento genético de la palma aceitera y producción de semillas en ASD Costa Rica. *ASD Oil Palm Papers*, 47, 2-18.
- Alvarado Gamboa, L. F. (2021). *Proyecciones de cambio climático regionalizadas para Costa Rica (Escenarios RCP-2.6 y RCP8.5)*. San José, Costa Rica: Ministerio del Ambiente Energía [MINAE]; Instituto Meteorológico Nacional [IMN].
- APSIM Initiative [AI]. (n.d.). Documentation: APSIM 7.10.
- Ayala D., I. M., Romero Angulo, H. M., Tupaz Vera, A. A., Daza, E. S., Rincón N., Á. H., & Caicedo Zambrano, A. F. (2017). *Comportamiento agronómico de cultivares comerciales de palma de aceite en Campo Experimental Palmar de la Vizcaína*. Bogotá, Colombia: Centro de Investigación en Palma de Aceite [Cenipalma].
- Barcelos, E., De Almeida Rios, S., Cunha, R. N. V., Lopes, R., Motoike, S. Y., Babiychuk, E., ... Kushnir, S. (2015). Oil palm natural diversity and the potential for yield improvement. *Frontiers in Plant Science*, 6(3), 1-16. <https://doi.org/10.3389/fpls.2015.00190>
- Billote, N. (2019). Desarrollo de herramientas genómicas para el mejoramiento de palma de aceite. El proyecto OPGP. *Revista Palmas*, 40(Especial T), 121-129.
- Bosi, C., Sentelhas, P. C., Pezzopane, J. R. M., & Santos, P. M. (2020). CROPGRO-Perennial Forage model parameterization for simulating Piatã palisade grass growth in monoculture and in a silvopastoral system. *Agricultural Systems*, 177, 102724. <https://doi.org/10.1016/j.agry.2019.102724>
- Carberry, P. S., Liang, W. -I., Twomlow, S., Holzworth, D. P., Dimes, J. P., McClelland, T., ... Keating, B. A. (2013). Scope for improved eco-efficiency varies among diverse cropping systems. *Proceedings of the National Academy of Sciences*, 110(21), 8381-8386. <https://doi.org/10.1073/pnas.1208050110>
- Carr, M. (2011). The water relations and irrigation requirements of oil palm (*Elaeis guineensis*): A review. *Experimental Agriculture*, 47(4), 629-652. <https://doi.org/10.1017/S0014479711000494>
- Chauhan, Y. S., Solomon, K. F., & Rodriguez, D. (2013). Characterization of north-eastern Australian environments using APSIM for increasing rainfed maize production. *Field*

Crops Research, 144, 245–255. <https://doi.org/10.1016/j.fcr.2013.01.018>

- Chen, J., Brissette, F. P., & Leconte, R. (2012). WeaGETS – a Matlab-based daily scale weather generator for generating precipitation and temperature. *Procedia Environmental Sciences*, 13(2011), 2222–2235. <https://doi.org/10.1016/j.proenv.2012.01.211>
- Ciaburro, G., & Venkateswaran, B. (2017). *Neural Networks with R: Smart models using CNN, RNN, deep learning, and artificial intelligence principles*. Birmingham, Mumbai: Packt Publishing Ltd.
- Corley, R. H. V., & Tinker, P. B. (2016). *The oil palm* (5th ed.). Hoboken, USA: Wiley-Blackwell.
- Culman, M., de Farias, C. M., Bayona, C., & Cabrera Cruz, J. D. (2019). Using agrometeorological data to assist irrigation management in oil palm crops: A decision support method and results from crop model simulation. *Agricultural Water Management*, 213, 1047–1062. <https://doi.org/10.1016/j.agwat.2018.09.052>
- Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random Forests. In C. Zhang & Y. Ma (Eds.), *Ensemble Machine Learning* (pp. 157–175). NY, USA: Springer Science+Business Media. https://doi.org/10.1007/978-1-4419-9326-7_5
- Deligios, P. A., Chergia, A. P., Sanna, G., Solinas, S., Todde, G., Narvarte, L., & Ledda, L. (2019). Climate change adaptation and water saving by innovative irrigation management applied on open field globe artichoke. *Science of The Total Environment*, 649, 461–472. <https://doi.org/10.1016/j.scitotenv.2018.08.349>
- Dudek, A. (2020). Silhouette index as clustering evaluation tool. In K. Jajuga, J. Batóg, & M. Walesiak (Eds.), *Classification and Data Analysis. SKAD 2019. Conference of the Section on Classification and Data Analysis of the Polish Statistical Association* (pp. 19–33). Springer, Cham. https://doi.org/10.1007/978-3-030-52348-0_2
- Easton, Z. M., & Bock, E. (2016). *Soil and soil water relationships*. Virginia Cooperative Extension.
- Escobar, R., Sterling, F., & Peralta, F. (2000). New ASD commercial oil palm seed materials. *Revista Palmas*, 21, 280–290.
- Feng, Y., Gong, D., Zhang, Q., Jiang, S., Zhao, L., & Cui, N. (2019). Evaluation of temperature-based machine learning and empirical models for predicting daily global solar radiation. *Energy Conversion and Management*, 198(May), 111780. <https://doi.org/10.1016/j.enconman.2019.111780>
- Foo, S. F. (1998). Impact of moisture on oil palm yield. *Kemajuan Penyelidikan* 32:5-17.
- Fraga, H., García de Cortázar Atauri, I., & Santos, J. . (2018). Viticultural irrigation demands under climate change scenarios in Portugal. *Agricultural Water Management*, 196, 66–74. <https://doi.org/10.1016/j.agwat.2017.10.023>
- Ghini, R., Bettioli, W., & Hamada, E. (2011). Diseases in tropical and plantation crops as affected by climate changes: current knowledge and perspectives. *Plant Pathology*, 60(1), 122–132. <https://doi.org/10.1111/j.1365-3059.2010.02403.x>

- Gomes, F. J., Bosi, C., Pedreira, B. C., Santos, P. M., & Pedreira, C. G. S. (2020). Parameterization of the APSIM model for simulating palisadegrass growth under continuous stocking in monoculture and in a silvopastoral system. *Agricultural Systems*, 184, 102876. <https://doi.org/10.1016/j.agry.2020.102876>
- Günther, F., & Fritsch, S. (2010). neuralnet: Training of Neural Networks. *The R Journal*, 2(1), 30–38.
- Guzmán Arias, I., & Calvo-Alvarado, J. C. (2012). Recursos hídricos de la Cuenca Alta del Río Tempisque, Costa Rica (Nota técnica). *Revista Tecnología En Marcha*, 25(4), 63. <https://doi.org/10.18845/tm.v25i4.620>
- Herrera, M., Natarajan, S., Coley, D. A., Kershaw, T., Ramallo-González, A. P., Eames, M., ... Wood, M. (2017). A review of current and future weather data for building simulation. *Building Services Engineering Research and Technology*, 38(5), 602–627. <https://doi.org/10.1177/0143624417705937>
- Hidalgo, H. G., Alfaro, E. J., & Quesada-Montano, B. (2017). Observed (1970–1999) climate variability in Central America using a high-resolution meteorological dataset with implication to climate change studies. *Climatic Change*, 141(1), 13–28. <https://doi.org/10.1007/s10584-016-1786-y>
- Hidalgo, Hugo G., & Alfaro, E. J. (2015). Skill of CMIP5 climate models in reproducing 20th century basic climate features in Central America. *International Journal of Climatology*, 35(12), 3397–3421. <https://doi.org/10.1002/joc.4216>
- Hochman, Z., Gobbett, D., Horan, H., & Navarro Garcia, J. (2016). Data rich yield gap analysis of wheat in Australia. *Field Crops Research*, 197, 97–106. <https://doi.org/10.1016/j.fcr.2016.08.017>
- Holzworth, D., Huth, N. I., Fainges, J., Brown, H., Zurcher, E., Cichota, R., ... Snow, V. (2018). APSIM Next Generation: Overcoming challenges in modernising a farming systems model. *Environmental Modelling & Software*, 103, 43–51. <https://doi.org/10.1016/j.envsoft.2018.02.002>
- Holzworth, D. P., Huth, N. I., DeVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., ... Keating, B. A. (2014). APSIM – Evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software*, 62, 327–350. <https://doi.org/10.1016/j.envsoft.2014.07.009>
- Hunt, J. R., Browne, C., McBeath, T. M., Verburg, K., Craig, S., & Whitbread, A. M. (2013). Summer fallow weed control and residue management impacts on winter crop yield though soil water and N accumulation in a winter-dominant, low rainfall region of southern Australia. *Crop and Pasture Science*, 64(9), 922–934. <https://doi.org/10.1071/CP13237>
- Huo, J., Shi, T., & Chang, J. (2016). Comparison of Random Forest and SVM for electrical short-term load forecast with different data sources. In *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)* (pp. 1–4). Beijing, China: IEEE. <https://doi.org/10.1109/ICSESS.2016.7883252>
- Huth, N.I., Henderson, C., & Peake, A. (2009). Development and testing of a horticultural

- crop model within APSIM. In R. S. Anderssen, R. D. Braddock, & L. T. H. Newham (Eds.), *Interfacing Modelling and Simulation with Mathematical and Computational Sciences. 18th World Imacs Congress and Modsim09 International Congress on Modelling and Simulation* (pp. 526–532). Cairns, Australia.
- Huth, Neil I., Banabas, M., Nelson, P. N., & Webb, M. (2014). Development of an oil palm cropping systems model: Lessons learned and future directions. *Environmental Modelling & Software*, *62*, 411–419. <https://doi.org/10.1016/j.envsoft.2014.06.021>
- Instituto Meteorológico de Costa Rica. (2008). Clima, variabilidad y cambio climático en Costa Rica, 75. Retrieved from http://www.cambioclimaticocr.com/multimedia/recursos/mod-1/Documentos/el_clima_variabilidad_y_cambio_climatico_en_cr_version_final.pdf
- Instituto Meteorológico Nacional [IMN]. (n.d.). *Pacífico Central*.
- Intara, Y. I., Nusantara, A. D., Supanjani, S., Caniago, Z., & Ekawita, R. (2018). Oil palm roots architecture in response to soil humidity. *International Journal of Oil Palm*, *1*(2), 79–89.
- Intergovernmental Panel on Climate Change [IPCC]. (2014). *Climate change 2014: Synthesis report. Contribution of working groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. (Core Writing Team, R. K. Pachauri, & L. A. Meyer, Eds.). Geneva, Switzerland: IPCC.
- Intergovernmental Panel on Climate Change [IPCC]. (2018). Annex I: Glossary. In V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, ... T. Waterfield (Eds.), *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change* (pp. 540–562). Geneva, Switzerland: IPCC.
- J. Chen, F. P. Brissette, R. Leconte, & A. Caron. (2012). A Versatile Weather Generator for Daily Precipitation and Temperature. *Transactions of the ASABE*, *55*(3), 895–906. <https://doi.org/10.13031/2013.41522>
- Karna, A., & Gibert, K. (2022). Automatic identification of the number of clusters in hierarchical clustering. *Neural Computing and Applications*, *34*(1), 119–134. <https://doi.org/10.1007/s00521-021-05873-3>
- Keating, B. ., Carberry, P. ., Hammer, G. ., Probert, M. ., Robertson, M. ., Holzworth, D., ... Smith, C. . (2003). An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, *18*(3–4), 267–288. [https://doi.org/10.1016/S1161-0301\(02\)00108-9](https://doi.org/10.1016/S1161-0301(02)00108-9)
- Kumar, V., Jain, M., Rani, V., Kumar, A., Kumar, S., & Naresh. (2018). A review of soil compaction - Concerns, causes and alleviation. *International Journal of Plant & Soil Science*, *22*(4), 1–9. <https://doi.org/10.9734/IJPSS/2018/40351>
- Kweku, D., Bismark, O., Maxwell, A., Desmond, K., Danso, K., Oti-Mensah, E., ... Adormaa, B. (2018). Greenhouse effect: Greenhouse gases and their impact on global warming. *Journal of Scientific Research and Reports*, *17*(6), 1–9.

<https://doi.org/10.9734/JSRR/2017/39630>

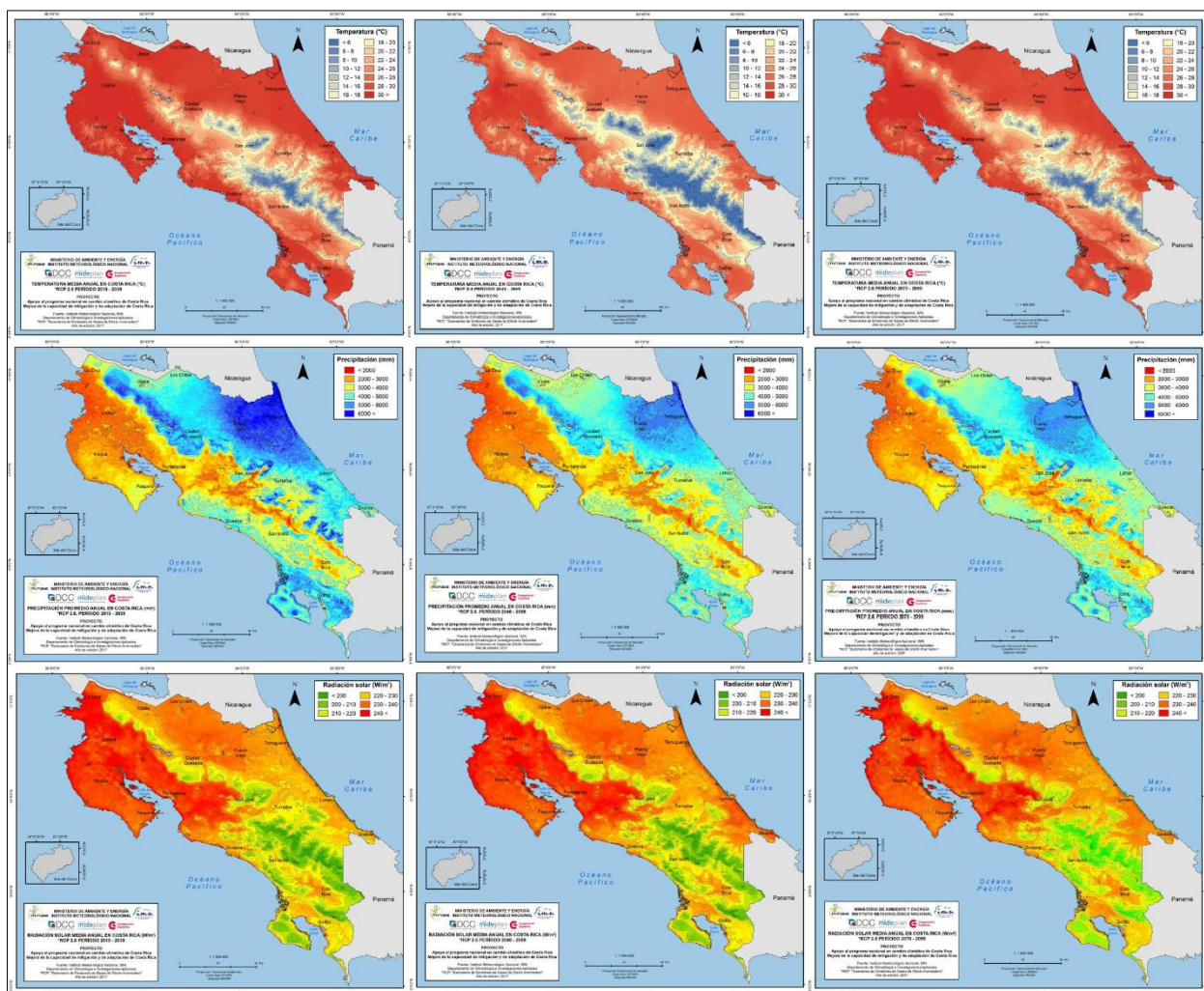
- Lee, K. Y., Kim, K. H., Kang, J. J., Choi, S. J., Im, Y. S., Lee, Y. D., & Lim, Y. S. (2017). Comparison and analysis of linear regression & artificial neural network. *International Journal of Applied Engineering Research*, 12(20), 9820–9825.
- Lilley, J. M., & Kirkegaard, J. A. (2016). Farming system context drives the value of deep wheat roots in semi-arid environments. *Journal of Experimental Botany*, 67(12), 3665–3681. <https://doi.org/10.1093/jxb/erw093>
- Lin, F., Chen, X., & Yao, H. (2017). Evaluating the use of Nash-Sutcliffe efficiency coefficient in goodness-of-fit measures for daily runoff simulation with SWAT. *Journal of Hydrologic Engineering*, 22(11), 05017023. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001580](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001580)
- Luo, Z., Wang, E., Baldock, J., & Xing, H. (2014). Potential soil organic carbon stock and its uncertainty under various cropping systems in Australian cropland. *Soil Research*, 52(5), 463–475. <https://doi.org/10.1071/SR13294>
- Mall, R. K., Gupta, A., & Sonkar, G. (2017). Effect of climate change on agricultural crops. In *Current Developments in Biotechnology and Bioengineering* (pp. 23–46). New Delhi, India: Elsevier. <https://doi.org/10.1016/B978-0-444-63661-4.00002-5>
- Meijaard, E., Garcia-Ulloa, J., Sheil, D., Wich, S. ., Carlson, K. ., Juffe-Bignoli, D., & Brooks, T. . (2018). *Oil palm and biodiversity. A situation analysis by the IUCN Oil Palm Task Force*. Gland, Switzerland: International Union for Conservation of Nature and Natural Resources [IUCN]. <https://doi.org/10.2305/IUCN.CH.2018.11.en>
- Messina, C. D., Podlich, D., Dong, Z., Samples, M., & Cooper, M. (2011). Yield–trait performance landscapes: from theory to application in breeding maize for drought tolerance. *Journal of Experimental Botany*, 62(3), 855–868. <https://doi.org/10.1093/jxb/erq329>
- Moriasi, D., Arnold, J., Van-Liew, M., Bingner, R., Harmel, R., & Veith, T. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE*, 50(3), 885–900. <https://doi.org/10.13031/2013.23153>
- Naciones Unidas. (2017). *Procesos de adaptación al cambio climático: análisis de América Latina*. Santiago, Chile.
- Okoro, S. U., Schickhoff, U., Boehner, J., Schneider, U. A., & Huth, N. I. (2017). Climate impacts on palm oil yields in the Nigerian Niger Delta. *European Journal of Agronomy*, 85, 38–50. <https://doi.org/10.1016/j.eja.2017.02.002>
- ORGANIZACIÓN DE LAS NACIONES UNIDAS PARA LA AGRICULTURA Y LA ALIMENTACIÓN. (2009). *Guía para la descripción de suelos. Organización de las Naciones Unidas para la Agricultura y la Alimentación* (Vol. 3).
- Paterson, R. R. M., Kumar, L., Taylor, S., & Lima, N. (2015). Future climate effects on suitability for growth of oil palms in Malaysia and Indonesia. *Scientific Reports*, 5(1), 14457. <https://doi.org/10.1038/srep14457>

- Paterson, R. R. M., & Lima, N. (2018). Climate change affecting oil palm agronomy, and oil palm cultivation increasing climate change, require amelioration. *Ecology and Evolution*, 8(1), 452–461. <https://doi.org/10.1002/ece3.3610>
- Pirker, J., Mosnier, A., Kraxner, F., Havlík, P., & Obersteiner, M. (2016). What are the limits to oil palm expansion? *Global Environmental Change*, 40, 73–81. <https://doi.org/10.1016/j.gloenvcha.2016.06.007>
- Rai, R. K., Singh, V. P., & Upadhyay, A. (2017). Chapter 17 - Soil Analysis. In *Planning and Evaluation of Irrigation Projects: Methods and Implementation* (pp. 505–523). Academic Press. <https://doi.org/10.1016/B978-0-12-811748-4.00017-0>.
- Rethinam, P., & Murugesan, P. (2018). Global perspective of germplasm and breeding for seed production in oil palm. *International Journal of Oil Palm*, 10(1–2), 17–34.
- Salamanca, A., & Siavosh, S. (2005). La densidad aparente y su relación con otras propiedades en suelos de la zona cafetera Colombiana. *Cenicafé*, 56(4), 381–397.
- Sarkar, M. S. K., Begum, R. A., & Pereira, J. J. (2020). Impacts of climate change on oil palm production in Malaysia. *Environmental Science and Pollution Research*, 27(9), 9760–9770. <https://doi.org/10.1007/s11356-020-07601-1>
- Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal: Promoting Communications on Statistics and Stata*, 20(1), 3–29. <https://doi.org/10.1177/1536867X20909688>
- Silva, P. Y. De, Fernando, C. N., Wijethunge, D. D., & Fernando, S. D. (2018). Recursive hierarchical clustering algorithm. *International Journal of Machine Learning and Computing*, 8(1), 1–7. <https://doi.org/10.18178/ijmlc.2018.8.1.654>
- Smith, M. R. (2021). Robust analysis of phylogenetic tree space. *Systematic Biology*, syab100. <https://doi.org/10.1093/sysbio/syab100>
- Solano Quintero, J., & Villalobos Flores, R. (2001). Aspectos fisiográficos aplicados a un bosquejo de regionalización geográfico climático de Costa Rica. *Revista Tópicos Meteorológicos y Oceanográficos*, 8(1), 26–39.
- Suresh, K. (2013). Adaptation and Mitigation Strategies for Climate-Resilient Oil Palm. In *Climate-Resilient Horticulture: Adaptation and Mitigation Strategies* (pp. 199–211). India: Springer India. https://doi.org/10.1007/978-81-322-0974-4_18
- Tiemann, T. T., Donough, C. R., Li, Y. L., Härdter, R., Norton, R., Tao, H. H., ... Oberthür, T. (2018). Chapter four - Feeding the palm: A review of oil palm nutrition. *Advances in Agronomy*, 152, 149–243. <https://doi.org/10.1016/bs.agron.2018.07.001>.
- Tseng, S., Chen, C., & Senarath, S. U. S. (2020). Evaluation of multi-site precipitation generators across scales. *International Journal of Climatology*, 40(10), 4622–4637. <https://doi.org/10.1002/joc.6480>
- Vogeler, I., Lilburne, L., Webb, T., Cichota, R., Sharp, J., Carrick, S., ... Snow, V. (2022). S-map parameters for APSIM. *MethodsX*, 101632. <https://doi.org/10.1016/j.mex.2022.101632>

- Wang, J., Wang, E., & Liu, D. L. (2011). Modelling the impacts of climate change on wheat yield and field water balance over the Murray–Darling Basin in Australia. *Theoretical and Applied Climatology*, 104(3–4), 285–300. <https://doi.org/10.1007/s00704-010-0343-2>
- Whish, J. P. M., Herrmann, N. I., White, N. A., Moore, A. D., & Kriticos, D. J. (2015). Integrating pest population models with biophysical crop models to better represent the farming system. *Environmental Modelling & Software*, 72, 418–425. <https://doi.org/10.1016/j.envsoft.2014.10.010>
- Woittiez, L. S., Wijk, M. T. Van, Slingerland, M., Noordwijk, M. Van, & Giller, K. E. (2017). Yield gaps in oil palm: A quantitative review of contributing factors. *European Journal of Agronomy*, 83, 57–77. <https://doi.org/10.1016/j.eja.2016.11.002>
- Woznicki, S. A., & Nejadhashemi, A. P. (2015). Journal of Hydrology : Regional Studies Climate change and irrigation demand: Uncertainty and adaptation. *Journal of Hydrology: Regional Studies*, 3, 247–264. <https://doi.org/10.1016/j.ejrh.2014.12.003>
- Wright Gilmore, J. (2006). *Estudio del potencial solar en Costa Rica*. San José, Costa Rica.
- Wright, J. (2009). Estudio del potencial solar en Costa Rica. *Uniciencia*, 23(1–2), 19–40.
- Xu, Q., Zhang, Q., Liu, J., & Luo, B. (2020). Efficient synthetical clustering validity indexes for hierarchical clustering. *Expert Systems with Applications*, 151, 113367. <https://doi.org/10.1016/j.eswa.2020.113367>
- Yang, L., Zhong, P., Zhu, F., Ma, Y., Wang, H., Li, J., & Xu, C. (2022). A comparison of the reproducibility of regional precipitation properties simulated respectively by weather generators and stochastic simulation methods. *Stochastic Environmental Research and Risk Assessment*, 36(2), 495–509. <https://doi.org/10.1007/s00477-021-02053-6>
- Yanyang, Y., Xu, J., Zhao, J., Li, N., & He, Y. (2017). Study on spring wheat yield change of inner Mongolia in time and space based on APSIM model. In *Proceedings of the 2016 6th International Conference on Advanced Design and Manufacturing Engineering (ICADME 2016)*. Paris, France: Atlantis Press. <https://doi.org/10.2991/icadme-16.2016.79>
- Zeybek, M. (2018). Nash-Sutcliffe efficiency approach for quality improvement. *Journal of Applied Mathematics and Computation (JAMC)*, 2(11), 496–503. <https://doi.org/10.26855/jamc.2018.11.001>

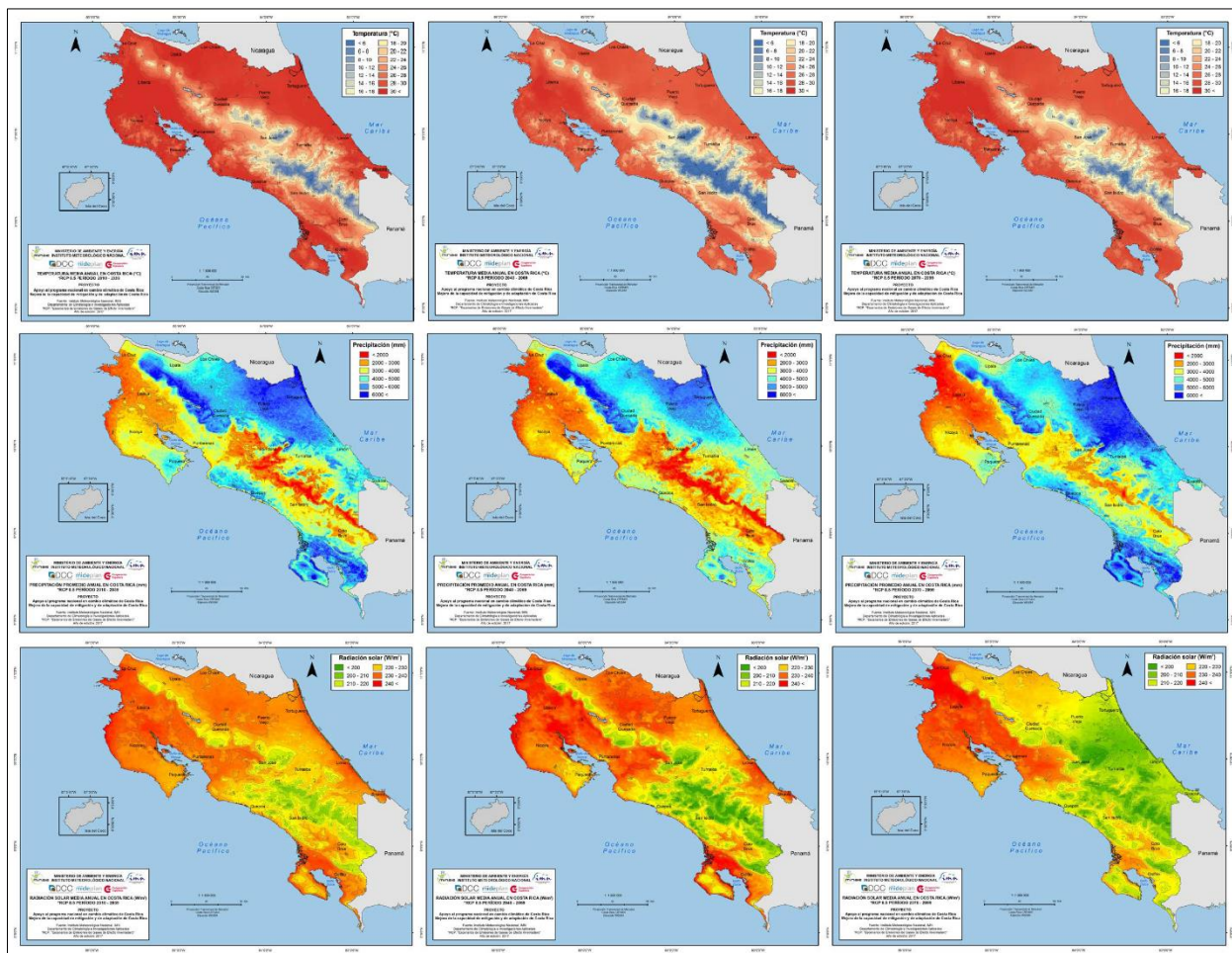
8. ANEXOS

Anexo 1 Proyecciones de temperatura, precipitación y radiación solar en Costa Rica en el escenario de RCP2.6 del 2010 al 2099.



Nota. Filas (izquierda a derecha): Temperatura media anual ($^{\circ}\text{C}$), Precipitación promedio anual (mm) y Radiación solar media anual (W/m^2). Columnas (arriba hacia abajo): corto plazo (2010-2039), mediano plazo (2040-2069) y largo plazo (2070-2099). Adaptado de *Proyecciones de cambio climático regionalizadas para Costa Rica (Escenarios RCP-2.6 y RCP8.5)* por Alvarado Gamboa, L. F., 2021, Ministerio del Ambiente Energía [MINAE], Instituto Meteorológico Nacional [IMN] (<http://cglobal.imn.ac.cr/index.php/publications/proyecciones-de-cambio-climatico-regionalizadas-para-costa-rica-escenarios-rcp-2-6-y-rcp-8-5/>). Derechos reservados.

Anexo 2 Proyecciones de temperatura, precipitación y radiación solar en Costa Rica en el escenario de RCP8.5 del 2010 al 2099.



Nota. Filas (izquierda a derecha): Temperatura media anual (°C), Precipitación promedio anual (mm) y Radiación solar media anual (W/m²). Columnas (arriba hacia abajo): corto plazo (2010-2039), mediano plazo (2040-2069) y largo plazo (2070-2099). Adaptado de *Proyecciones de cambio climático regionalizadas para Costa Rica (Escenarios RCP-2.6 y RCP8.5)* por Alvarado Gamboa, L. F, 2021, Ministerio del Ambiente Energía [MINAE], Instituto Meteorológico Nacional [IMN] (<http://cglobal.imn.ac.cr/index.php/publications/proyecciones-de-cambio-climatico-regionalizadas-para-costa-rica-escenarios-rcp-2-6-y-rcp-8-5>). Derechos reservados.