Forecasting of yields in Jarillo peach crops at the Province of Pamplona using random variables

Pronóstico de rendimiento en cultivos de duraznero Jarillo en la Provincia de Pamplona utilizando variables aleatorias

DOI: http://doi.org/10.17981/ingecuc.18.1.2022.06

Artículo de Investigación Científica. Fecha de Recepción: 10/06/2021. Fecha de Aceptación: 10/02/2022.

Carlos Arturo Parra Ortega
Universidad de Pamplona. Pamplona (Colombia)
carapa@unipamplona.edu.co

Enrique Quevedo García
Universidad de Pamplona. Pamplona (Colombia)
enriquegarcia@unipamplona.edu.co

César Villamizar Quiñones
Universidad de Pamplona. Pamplona (Colombia)
csrvillamizar@unipamplona.edu.co

To cite this paper:
C. Parra Ortega, E. Quevedo García & C. Villamizar Quiñines, "Forecasting of yields in Jarillo peach crops at the Province of Pamplona using random variables", *INGE CUC*, vol. 18, no. 1, pp. 66–82, 2022. DOI: http://doi.org/10.17981/ingecuc.18.1.2022.06

Resumen

Introducción— Como resultado de un proyecto de investigación, en este artículo se muestra el pronóstico de rendimiento de cultivo de durazno variedad jarillo. Para ello se diseñó un modelo para simular la producción de frutos de durazno generando variables aleatorias de cantidad de frutos y peso total, a partir de distribuciones de probabilidad deducidas a partir de muestras de cultivos.

Objetivo— Pronosticar el rendimiento de un cultivo de durazno, mediante la simulación de variables que siguen una distribución de probabilidad asociadas al mismo, obteniendo un comportamiento estadístico similar a un escenario de producción real.

Metodología—Se hizo revisión bibliográfica de estudios sobre pronóstico de producción en otras especies vegetales. También se tomaron muestras de producción en haciendas de diversos pisos térmicos y se hizo un análisis de regresión lineal a intervalo fijo (stepwise) teniendo como variable dependiente el rendimiento y como variable independiente las dimensiones físicas de la rama. Además, se tomaron datos sobre la distribución de probabilidad de producción y con base en ella se diseñó e implementó un software simulador, con el cual se hicieron diversas simulaciones de escenarios de producción.

Resultados— Se obtuvieron modelos con menor número de variables resultantes de aplicar el procedimiento "stepwise" para pronosticar el número de frutos y rendimiento. Al caracterizar variables de entrada, se pasó a construir el modelo matemático con entradas aleatorias para pronosticar el rendimiento, tales como el área del cultivo, sistema de siembra, densidad de siembra, edad del cultivo y longitud de rama, área foliar, diámetros del fruto entre otras variables.

Conclusiones— Se logró demostrar que es factible pronosticar la el rendimiento de cultivo de duraznos en varios supuestos, a partir de muestras observadas en escenarios de producción real. Se logró implementar un modelo de pronóstico basado en variables aleatorias, cuya variabilidad con respecto a datos reales es significativamente pequeña.

Palabras clave— Producción de durazno; modelos estocásticos; simulación de variable discreta; sistemas de producción agrícola; sistribución de probabilidad

Abstract

Introduction— As a result of a research project, this article shows the crop yield forecast for a peach variety named jarillo. For this, a model was designed to simulate the production of peach fruits, generating random variables of the number of fruits and total weight, from probability distributions deduced from crop samples.

Objective— Forecasting the yield of a peach crop, by simulating variables that follow a probability distribution associated with it, obtaining a statistical behavior similar to a real production scenario.

Methodology— A bibliographic review was made of studies on production forecasting in other plant species. Production samples were also taken from farms in various zones and a linear regression analysis at a fixed interval (stepwise) was made, taking yield as a dependent variable and the physical dimensions of the branch as an independent variable. In addition, data was collected for determine the production probability distribution and based on it a simulator software was designed and implemented, with which various simulations of production scenarios were made.

Results— Models were obtained with a lower number of variables resulting from applying the stepwise procedure in order to forecasting the number of fruits and performance. When characterizing input variables, the mathematical model was built with random inputs to predict yield, such as crop area, planting system, planting density, crop age and branch length, leaf area, fruit diameters, among others variables.

Conclusions— Is feasible the forecasting of peach crops yield under several assumptions, from samples observed in real production scenarios. It was posible to implement a forecast model based on random variables, whose variability with respect to real data is significantly small.

Keywords— Agricultural production systems; Stochastical models; Peach production; Discrete-variable simulation; Probability distribution

INGE CUC vol. 18 no. 1, pp. 66-82. January - June, 2022
Barranquilla. ISSN 0122-6517 Impreso, ISSN 2382-4700 Online

I. Introduction

A. *Background*

The peach (Prunus persica) is considered one of the most important stone fruits of the Rosaceae family worldwide, after the apple and the pear. In 2019, approximately 25.74 million tons of peaches were produced, mainly in temperate zones, where China was the first producer with 61.5% [1]. This fruit is located in the top of the 10 most produced worldwide [2]. However, currently the crop of peach has been extended to non-traditional areas in subtropical and tropical regions, where the climate differs from its natural habitat, and its consumption is also increasing [3].

In the province of Pamplona, south of the department of Norte de Santander (Colombia), peach crop is considered an important economic activity as it is a relevant source of income for producers in this region [4]. The crop of the Jarillo variety with low cold requirements is of Venezuelan origin and is practiced in the department of Norte de Santander [5]-[6]. In this region, the harvested area by varieties of Venezuelan origin has been increasing exponentially between 2007 and 2019 in the department of Norte de Santander, ranging from 167 ha to 835.7 ha. It should be noted that peach production in the province of Pamplona is carried out by small farmers, who earn their income from this production system. In 2019, 30 232 tons of peach were produced in the high Colombian tropics, with a yield of 9.96 ton/ha and a harvest area of 2 077 ha [7]. Although the department of Boyacá stands out as the quintessential producer of peaches with a production of 13 802 tons and yields of 10.6 ton/ha and participates with 45.65% in national production, the department of Norte de Santander ranks second with a production of 11,246 tons and yields of 12.25 ton/ha [7]. Regarding this yields, are low, compared with the world average of 16.85 ton/ha [1]. The production systems of this variety turn out to be economically efficient (average profit margin 59%) and energetically deficient (average energy efficiency of 0.58) [8]. All farms planted with the variety use agrochemicals to fertilize, combat pests and diseases and induce flowering, in the style of the green revolution [9].

In the production of peaches for fresh consumption, the size of the fruit is one of the determining factors of quality, since it is a fundamental requirement demanded by the consumer [10]. Regarding the invention processes patented in recent years for Colombia in peach, it was established that the developments are meager, registering some processes in the field of preserving food or new products [11].

B. *Statistical models to predict the productive behavior of deciduous fruit trees at different altitude levels*

A crop mathematical model is designed for simulating the dynamic of crop growth and yielding, by the use of numerical integration of logical processes with the aid of computer programs. More specifically this implies a program which describes the dynamics of crop growth in relation to the environment, operating in time steps whose magnitude are less than the growing period and with the ability to obtain variables that describe the state of the crop at different timepoints, for example: biomass per unit area, stage of development, yield, leaf nitrogen content, etc. The end users of this mathematical models classify them into those used as tools in: research, decision making, education, training or technology transfer [12]-[13]. There are simulation models for a crop, for example CERES-Wheat, CERES-Maize, CORNGRO, SIMTAG, SIMBA for wheat and corn, or CROPGRO-SOYBEAN, WOFOST, and SOYSIM for soybeans. SUCROS-CACAO, CROPGRO-DRY BEAN, CROPGRO-Tomato [14], or on the other hand, can be generic. Generic models can be applied to various species by using crop-specific parameters: some of these models are DSSAT-CSM, DAISY, SOILN, EPIC, WOFOST, CROPSYST, APSIM, AQUACROP [15], and STICS. they allow the user to organize and manipulate information on crops, soils and meteorological data, being able to execute the crop models in different ways and analyze their results, due to the fact that they share a common input and output file format [16], [12]. For EU/IICA [17] and Netherlands [18], two types of approach can be carried out in the construction of crop models: one descriptive and the other explanatory (mechanistic). According to the UNI [13], the models can be categorized into empirical, mechanistic or causal, functional or process-oriented models.

Empirical models are reduced to a simple mathematical expression that transforms a set of "input" variables into an "output" of results, without even attempting to describe the process under study. Regression models and many statistical models are good examples of empirical predictive models. Empirical models are generally simple, require less data than causal models, and are easier to build and use. But they have the disadvantage of not being able to be improved or extended beyond the conditions in which they were made. Therefore, they can easily be applied incorrectly and, furthermore, lead their users without critical capacity to much confusion in the interpretation of the processes under study.

Causal or "mechanistic" models have the purpose of describing in the most exact and verifiable way possible, the physical, chemical and biological processes involved in the soil-plant relationship from the available input data. This type of models can be used to predict responses to fertilization procedures, evaluate the effect of environmental change, coordinate and structure research, its main disadvantages being that it requires larger volumes of data and extensive preliminary activities for its calibration and validation. The functional ones are used to explain processes such as those shown with Hart's models, to explain the components and their interrelationships in agricultural production systems [19]. After calibration, for example, a model such as Yield-SAFE provides credible descriptions of measured crop and tree (Populus spp) yields in single crop systems and forestry capacity at the two sites. The models are calibrated [20].

C. *Random variables*

A random variable generator is a mathematical tool that allows statistical modeling of processes of various kinds, in particular, generating a sample of variables that can be used as a production forecast. The main difficulty lies in two aspects, documented [21]-[24]. One of them is to determine the probability density function of a variable (pdf, from here on), given its historical behavior, and after obtaining this function, obtain a mathematical algorithm that allows its value to be generated in such a way that a large set of these values behave statistically similar to that of the corresponding pdf [25]. In Fig. 1 are shown the theoretical graph of the normal distribution, where the value of a variable is shown on the horizontal axis, and the relative frequency (probability) of that value is shown on the vertical axis. Adding to this, are showed a distribution obtained by generating 900 values and placing them in a histogram using the Box-Müller algorithm. The further the values are from the mean, the probability for them decreases.

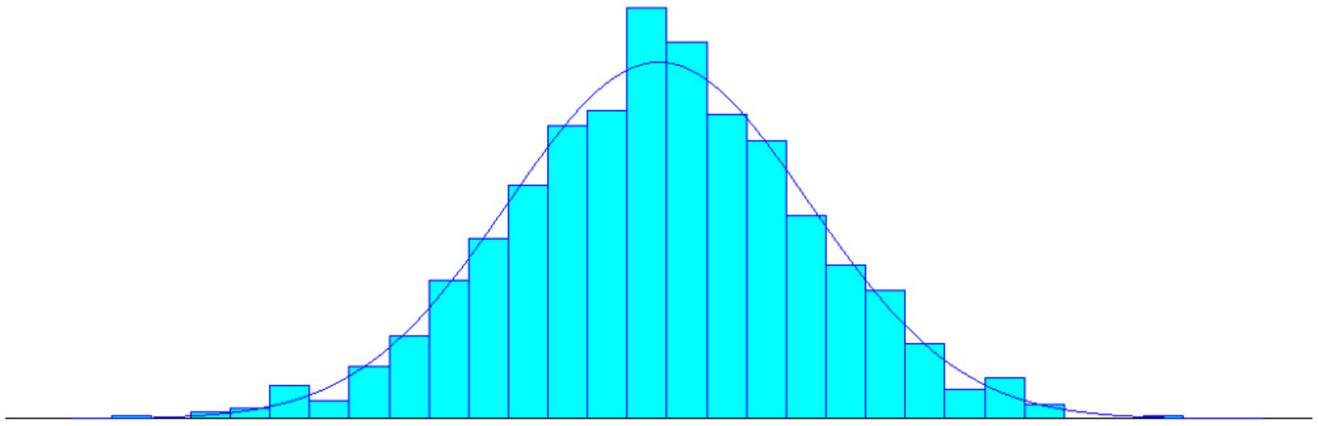


Fig. 1. The normal distribution, theoretical and sampled. Where the horizontal axis shows the value of a variable, and the vertical axis shows the relative frequency (probability) of such value. The theoretical value is in a solid line, while the sample appears as a bar in each interval.
Source: Authors.

This work seeks to study the case of the agronomic production of peaches of the Jarillo variety, trying to find a statistical distribution of the variables associated with yield, given some physical characteristics of the plant, in order to find a behavior that can adjust to a known distribution and then proceed to generate random variables and thus forecast their long-term production. For the case described in this study, it is about generating a yield forecast for various farms in the province of Pamplona, where the probability of production of each plant is addressed, and according to crop conditions, establish the probable value for its performance.

II. Methodology

A. *Vegetal material and featuring of the study áre*

The Jarillo variety under study has the following agronomic characteristics: it has yellow shell and pulp, small size, rounded shape, high prolific production and susceptible to post-harvest handling. This peach has established itself in small areas. The plants are large due to the established planting distance of 6 m to 7 m between plants and furrows, distributed according to homogeneous slope conditions and agronomic management, forming plots with peach trees in full agronomic production from sexual seed and with an age older than ten years [26].

This research was carried out within the physiographic zone of the Santanderean mountain, which included the middle zones of the eastern mountain range in the department of Norte de Santander, between 1 600 m and 2 300 m above sea level. In Pamplona, "Delicias" farm, village of Chíchira (7° 22' 43.6" N and 72° 37'41.1" W). The altitude was 2 170 masl, average temperature 16°C, rainfall 933.9 mm, bimodal regime, and the soils belong to the Inceptisol order. In Chitagá, "Recuerdo" farm, Carrillo village (7° 11' 15" N and 72° 39' 7.3" W), whose altitude is 1 870 meters above sea level, with an average temperature of 18°C, precipitation of 879.5 mm, unimodal regime, and soils of the Inceptisol order. And in Pamplonita, "Bella Vista" farm, Batagá district (7° 26' 18.1" N and 72° 38' 9" W), with an altitude of 1 670 masl, average temperature 20°C, with 1200 mm of rain in bimodal regime and soils belonging to the Inceptisol order [27].

B. *Sampling*

Sampling was established by location, with the same slope, age of the trees, sexual origin, stage of development achieved with forced management and the same type of pruning and thinning during the production cycle. From the trees, 3 branches per tree were taken, which are the ones with the highest number of reproductive buds, 3 fruits of similar appearance in color and size, harvested from the middle part of the mixed branch, 3 leaves, located in the middle third of the tree in the part exposed to the sun and mixed branches of the middle third, between two secondary stages of development 72: green ovary surrounded by a crown of dying petals, the secondary stage 75: fruit approximately half the final size and in the secondary stage 87: fruit ripe for harvest [28]-[30].

C. *Morphophysiological variables*

A data matrix was involved in the study, for which the following variables were available: length (cm), width (cm), the product of length by width of the blade (cm^2), leaf area (cm^2), number of active leaves, length of the mixed branch (cm), branch perimeter (cm), average diameter of the branch (cm), equatorial and polar diameter of the fruit (cm), quotient between the equatorial and longitudinal diameter of the fruits, the fresh weight of the fruit (g) and the associated dry weight (g). The leaf blade area was determined in 108 mature leaves with approximately 2.5 months of age for the three farms (39 per farm), which were sampled during 22 samplings from stage 72 to harvest phenological stage (87), by means of the use of the ImageJ public domain program which performs image analysis [31]. The process began with the download of the photo (jpg extension) in the program and the inclusion of the limb length (cm) as input parameter, which was previously measured with an Ubermann electronic vernier caliper (Fa. Sodimac, Santiago , Chili). The measurements of the diameters were collected with the same vernier caliper. The dry matter (g) was measured after the fruits were dried, using a Mettler Toledo® dry hot air forced circulation electric drying oven. The mass was obtained using an electronic balance with two-digit Lexus® approximation with a measurement range of 3 600 g and precision: 0.01 g, it was obtained after 48 hours at 60°C in the oven and taking into account that the weight is constant. This situation is achieved by weighing at regular intervals until a constant weight is achieved. With the same scale, the fresh mass of the fruit was obtained. The number of active leaves was measured during 22 samplings in each of the farms. Measurements of the number of active leaves were made on the three randomly selected branches of each of the 39 trees.

D. *Productive behavior of the Jarillo peach variety in the province of Pamplona*

Two harvests (with their respective passes) of the fruits in the secondary stage 87 were collected for 39 trees located in the three thermal floors in the first cycle (2014), second productive cycle (2015). A digital H-110 scale with a capacity of 50 kg, a resolution of 20 g and a tolerance of 40 g made in China and paper bags were used to identify the production by tree and plot. The following variables were measured: number of fruits, total harvest per tree (kg/tree), fresh weight of the fruit (g), roundness index, longitudinal diameters (cm) and equatorial diameters (cm).

E. *Statistical analysis*

The best multiple linear regression models were chosen, using the SPPS 25 statistical package [51], with the Stepwise variable selection method, where the independent variables and the responses were introduced to estimate the fresh weight of the fruit, the total soluble solids and the total harvest per thermal floor taking into account that the R2 was at least 0.8, which is the statistic that represents the proportion of variation explained by the regression; the adjusted R2, the significance of the model in its anova test, the sum of square error, the mean square of the experimental error and the significance of the regression parameters. Regarding the validation of the regression models, the following assumptions were taken into account, which are: linearity, independence, homoscedasticity, normality and non-collinearity. Finally, a comparison of the observed data with the estimated data was made, to observe the relationship between them.

F. *Principal component analysis (PCA)*

The multivariate analysis consisted of the work of principal component tests; These tests were carried out with the statistical program R® Project for Statistical Computing [52] and the results were corroborated using the statistical package SPSS® v. 25.0 [51] and the Input Analyzer software v. 16.0 [53]. The results were supported for discussion in the relevant literature. Multivariate analysis was performed with the variables that most correlated with the performance and quality components; To find the minimum number of factors, the Kaiser rule was used, with which the eigenvalues of the correlation matrix R were compared, and the number of eigenvalues greater than unity was taken as the number of factors; Therefore, this criterion is an allusion to the analysis of principal components and it has been verified in simulations that, generally, it tends to underestimate the number of factors so its use is recommended to establish a lower limit, an upper limit would be calculated by applying this same criterion taking 1 as limit. According to American researchers [32] it is indicate that PCA is a multivariate technique widely used for the reduction of dimensions to a small set of uncorrelated variables called CP (Principal Component). A small set of uncorrelated variables is much easier to understand and use in PCA than a larger set of correlated variables. The first CP explains the largest possible variability in the data, and each subsequent component explains the largest possible amount of the subsequent variability [33].

G. *Multiple linear regression model*

The multiple regression model used to estimate the agronomic yield variables of the Jarillo variety peach tree was the following [34]-[35] (1):

$$Yi = \beta 0 + \beta 1\, X1i + \beta 2\, X2i + \beta 3\, X3i + \beta k\, Xki + \in i \quad (1)$$

$B0$ is the ordinate to the origin or intercept with the Y axis where $\beta 1$...
βk are the regression coefficients of the predictor
variable in the multiple equation.
$X1i$.... Xki are explanatory, predictor or regression variables,
and $\in i$ is the magnitude of a random variable with mean zero (0) and variance $\sigma 2$.

As in simple linear regression, the ß coefficients will indicate the increase in fresh weight, number of fruits and yield, the unit increase of the corresponding explanatory variable. Therefore, these coefficients will have the corresponding units of measure. Applying the method of main components [36] and linear regression, values were obtained for the morphological variables, allowing the characterization of peach tree populations in a wide range of distribution, such as an altitudinal gradient. Similar analyzes have been done on other crops such as potatoes [37].

H. *Design of simulator software to compute physiological production of prunus persica*

With the previous procedures the variables were featured and a mathematical procedure was designed, implemented with a software that can meet the requirements of the physiological modeling of peach (*P. persica*) production, exposed in the document of the statistical model. One way to organize it is to conceive of the system as interconnected components, in which each component performs some tasks. This organization is called the "system architecture". There are packages such as L-PEACH used in peach cultivation that predict the growth and distribution of photoassimilates in the plant [38]-[39]. Several simulation runs were executed to analyse the consistency of the simulated results with those initially observed, to determine the validity of such results.

III. Analysis and results

As a result of the aforementioned components and regression analysis, peach populations could be characterized, evaluated and compared based on morphophysiological descriptors and yield components related to the leaf and fruit. Table 1 shows that the variables that most contributed to the construction of the first component, which is the one that retains the greatest variability of the data, were the roundness index of the fruit, the total fresh weight of the fruit, the total dry weight of the fruit, the equatorial diameter of the fruit, the longitudinal diameter of the fruit, the length by width of the leaf and the leaf area. Very negative values of the first axis were associated with high values of the fruit and leaf measurements. The variables that contributed to the construction of the second index were those associated with the leaf and the fruit, as has been shown in other works [40]. In order to establish the variables that would contribute to environmental variability, quality and its contributions to the morphophysiological variables in the three main components were presented in Table 1.

Table 1.
Quality of representation of the variables in the components created through the PCA of the morphophysiological variables at year 2014 cycle.

Variables/ Quality of representation of the variables	Component 1	Component 2	Component 3	Remainder
Fruit roundness index	86.41	2.02	0.05	11.52
Total fresh weight of the fruit	87.71	3.49	0.34	8.46
Total dry weight of the fruit	83.34	7.07	0.06	9.53
Equatorial diameter of the fruit	71.3	10.88	1	16.82
Longitudinal diameter of the fruit	64.61	9.53	4.29	21.57
Branch length	0.59	4.88	62.94	31.59
Average branch diameter	8.98	32.57	22.77	35.68
Number of active leafs	1.81	8.98	60.15	29.06
Leaf width	0.21	69.83	3.08	26.88
Leaf length	16.56	76.39	1.1	5.95
Product of leaf Length and leaf Width	54.84	28.81	0.03	16.32
Leaf area	53.73	29.64	0.01	16.62

Source: Authors.

The variables that showed the greatest variance were those associated with the fruit, the leaf and the branch for the three heights or thermal floors as shown in Table 2.

Table 2.
Representation quality of variables in components created using the ACP of the morphophysiological variables at year 2015 cycle.

Variables/ Components	Component 1	Component 2	Component 3	Other
Fruit roundness index	79.37	12.14	0.16	8.33
Total fresh weight of the fruit	85.55	8.43	0.16	5.86
Total dry weight of the fruit	80.84	9.59	0.07	9.5
Equatorial diameter of the fruit	69.32	19.82	0.47	10.39
Longitudinal diameter of the fruit	68.71	17.78	0.08	13.43
Branch length	0.19	10.6	61.31	27.9
Average branch diameter	4.79	71.44	7.26	16.51
Average branch perimeter	3.69	56.96	14.32	25.03
Number of active leafs	3.81	23.43	30.97	41.79
Leaf width	4.02	75.91	6.26	13.81
Leaf length	21.95	63.85	4.81	9.39
Product of leaf Length and leaf Width	81.12	10.76	0.18	7.94
Leaf area	81.6	10.62	0.22	7.56

Source: Authors.

Table 3 shows the estimates of the regression parameters in each period taking into account the previous variables analyzed in the main components, thus, for example, in the first row there is the estimate of a multiple linear regression model such as has been done in other crops [41]-[42], for production on the hundredth day after defoliation, the multiple linear regression model in that period is (2):

$$\begin{aligned} y = \beta_0 - 674.49 + 36.11x_1 + 201.87x_2 \\ + \cdots - 23.62x_{11} - 5.25x_{12} \end{aligned} \qquad (2)$$

Where Y is the production in kilograms,
the explanatory variables (X) are found in the header of Table 3.

Next, are shown the results of models with the least number of variables resulting from applying the "stepwise" procedure to keep the least number of variables necessary to obtain a model that is as suitable as possible. In the different models elaborated at each time, it was observed that not all the variables are significant to explain production, either because of the little linear relationship that exists between this variable and production or because of the multicollinearity that exists between the variables. The stepwise procedure was performed to obtain a regression model using the necessary variables to obtain an adequate model for each period.

Table 3 shows the estimates of the parameters for the variables according to the days after defoliation. The cells where NA appears mean that in the period corresponding to the row the variable of the respective column is not statistically significant to include it in the model.

Table 3.
Parameters by time period.

Days after defoliation	Intercept	Leaf width	Leaf area	Ecuatorial diameter of the fruit	Longitudinal diameter of the fruit	Diamteter of the branch	Fruit roundness index	Leaf length	Branch length	Leaf width x length	Number of active leafs	Total fresh weigth of the fruit	Total dry weigth of the fruit
100	–468.31	16.15	–213.51	103.14	NA	NA	329.85	NA	NA	156.13	NA	–24.73	NA
101	–774.64	21.79	–215.98	120.29	NA	NA	347.55	NA	NA	163.14	NA	–26.11	NA
102	–854.53	NA	–241.20	NA	102.71	NA	468.43	–23.48	NA	189.07	NA	–27.28	NA
.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.
156	–1016.66	172.73	29.30	NA	NA	NA	339.74	213.84	NA	–60.95	NA	–9.25	NA
157	**–987.79**	**176.10**	**26.50**	**NA**	**NA**	**NA**	**323.33**	**219.74**	**NA**	**–60.95**	**NA**	**–9.06**	**NA**

Source: Authors.

The multiple determination coefficient shown in Table 4 establishes that the percentage of variability is explained by the multiple regression model, this is a value ranged from zero to one, and measures how adequate the linear regression model is to make predictions. On the other hand, the adjusted correlation coefficient is framed in the same previous concept, but it penalizes the use of a high number of variables in the model, that is, it does not only take into account the adjustment but does not consider the amount of information available for the elaboration of the models.

Column two shows the coefficient of determination of the model including all available variables, the respective adjusted coefficient of determination is presented in column three. For columns four and five the same measurements are presented, but for the reduced model. The variables associated with the fruit are of great importance to estimate crop yield [40].

Table 4.
Coefficient of determination.

Days after defoliation	Squared R of complete model	Adjusted squared R of complete model	Squared R of reduced model	Adjusted squared R of reduced model
100	0.69	0.37	0.67	0.57
101	0.69	0.38	0.68	0.58
102	0.69	0.37	0.68	0.58
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
156	0.67	0.34	0.63	0.51
157	0.66	0.32	0.62	0.50

Source: Authors.

Similarly, with the total number of fruits produced by each of the selected trees in the study farms was totalized the number of fruits produced in each tree by passing. It has been shown that the number of harvested fruits is the most important primary yield component to estimate the Jarillo peach harvest, a result similar to that obtained in other crops by [43].

It was proposed to develop a statistical model to predict the number of fruits and the production in kilograms of a tree given certain characteristics that can be measured in a given time. In this regard, the following considerations must be taken into account:

There are variables associated with the fruit, the mixed branch, the oblong-lancellolate leaf, and the average weight of a fruit in a Jarillo peach tree can be traced.

The variables that are available were measured over time (taking the days after defoliation as a reference) arranged in a design of repeated models, however, the time in which the information of the different variables is collected varies from farm to farm. In addition, is not the same in all the variables due to the influence of several factors, such as the edaphoclimatic ones.

The two variables to be modeled (dependent or response variables) are the number of fruits produced per tree and the production in kilograms of fruit per tree. These two variables are measured once at the end of the production cycle.

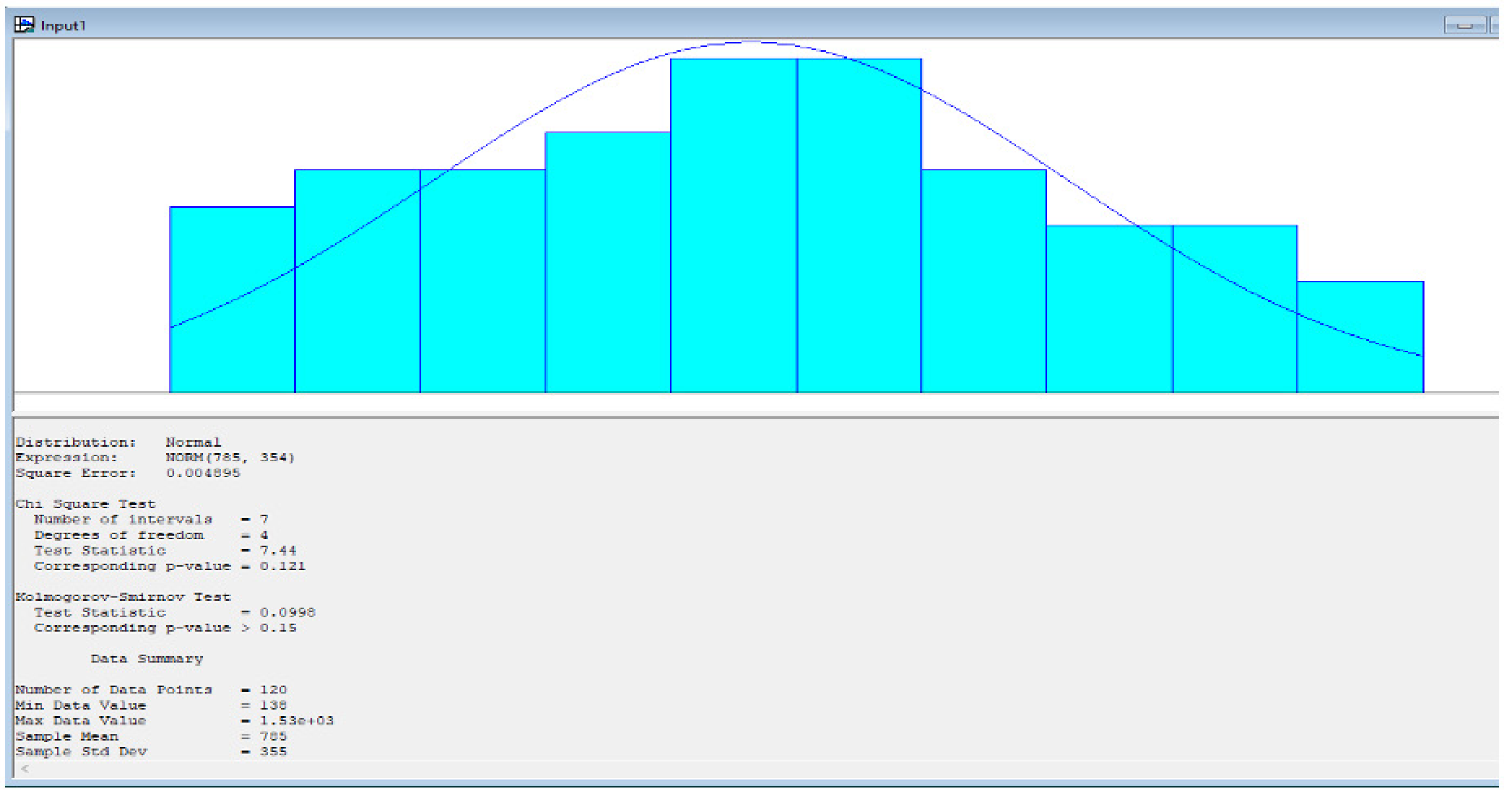
It is desirable to have a model that allows forecasting how much production will be at a given time (a given number of days after defoliation). In the production of peaches for fresh consumption, the size of the fruit is one of the determining factors of quality, since it is a fundamental requirement demanded by the consumer [44]-[46]. The peach fruits that are considered to be of good size are those greater than 59 mm in diameter. On the other hand, standardizing or classifying the product by size is one of the usual practices of preparation for the fresh market [47].

Fruit weight is an important hereditary quantitative factor of yield, fruit quality and consumer acceptability as stated in other work [47] and was best expressed by the influence of the environment on the physiology of trees in the lower altitudes.

Two technical limitations make it difficult to obtain a direct model for estimating production, first of all, how asynchronous are the measurements in the different variables and in the different farms, this is due to logistical and cost issues, and due to differences in the development of the plants of each of the farms.

Another limitation that exists is the aggregation of the variables of mixed branches, leaves and fruits through sample means, naturally due to costs and practical issues it would be unfeasible to measure each branch, fruit and leaf of a tree (in this case there would be no sampling error for each selected tree). However, having samples of these observation units (branch, fruit, leaf) generates a sampling error which is not quantifiable due to the auto-correlation effect caused by the same type of design used for repeated measures. In addition to the difficulties mentioned above, the production and number of fruits depends largely on various environmental, economic, ecological factors, among others, that are not easy to consider.

A. *Behavior of the probability distribution of the number of fruits*



```
Distribution:   Normal
Expression:     NORM(785, 354)
Square Error:   0.004895

Chi Square Test
  Number of intervals   = 7
  Degrees of freedom    = 4
  Test Statistic        = 7.44
  Corresponding p-value = 0.121

Kolmogorov-Smirnov Test
  Test Statistic        = 0.0998
  Corresponding p-value > 0.15

        Data Summary

Number of Data Points   = 120
Min Data Value          = 138
Max Data Value          = 1.53e+03
Sample Mean             = 785
Sample Std Dev          = 355
```

Fig. 2. Analysis of the number of fruits in a single harvest from the sample.
Source: Authors.

Making an analysis of the 39 plants from which production data was taken, it is found that in a single harvest the minimum number of fruits obtained in a tree was 138 and the maximum 1 530, being the average number of fruits per plant of this variety is 785, with a standard deviation of 354. Using the chi-square test, it is indicated that with a certainty level of 90% there are no reasons to reject the sample. Likewise, the square error of the goodness of fit test is 0.004895, which indicates that the sample taken fits this distribution and its parameters represent the sample. The Kolmogorov-Smirnov test also gives a good approximation value of 0.0998, which also indicates that there is no reason to reject the data. The distribution assumed for this sample is Normal, with the following parameters: mean 785 and standard deviation 354. It should be noted that there may be more than one harvest in a year. The Fig. 2 shows the analysis made using the input analyzer software. The frequency is observed in the 10 intervals analyzed and the theoretical normal distribution is shown as a continuous line.

B. *Behavior of the probability distribution of fruit weight*

The weight of the peach during each harvest is related to the number of fruits. However, the variability is high since a high number of fruits can have different sizes and does not imply a high final weight. Fig. 3 shows the dispersion diagram of the amount of fruit and the harvested weight in blue, compared to the regression of the data, in orange line. As can be seen, there is a relationship between the two variables, but with high variability, which suggests that they can be treated as two independent variables when simulating. Likewise, there is a greater number of observations near the average number of fruits, between 700 and 850 fruits, respectively.

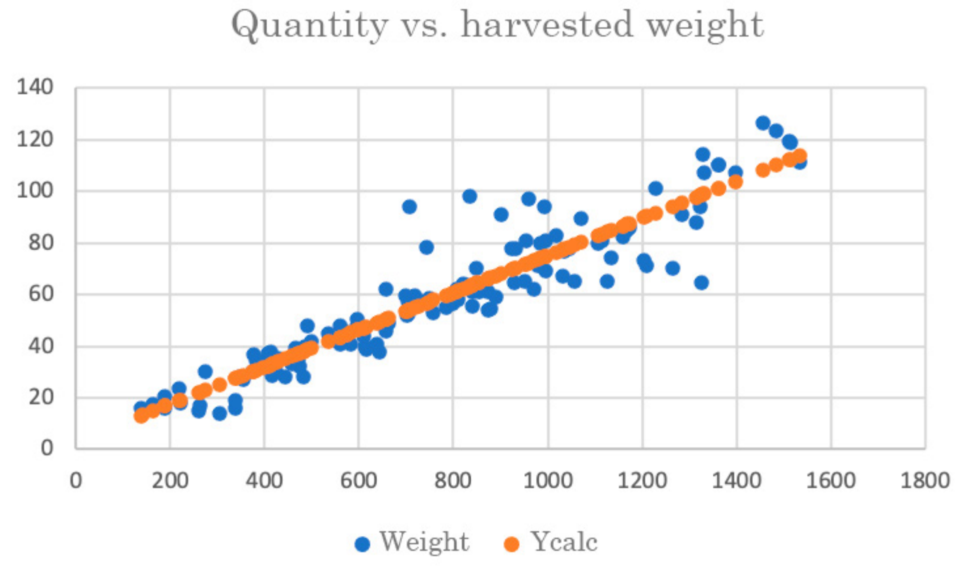
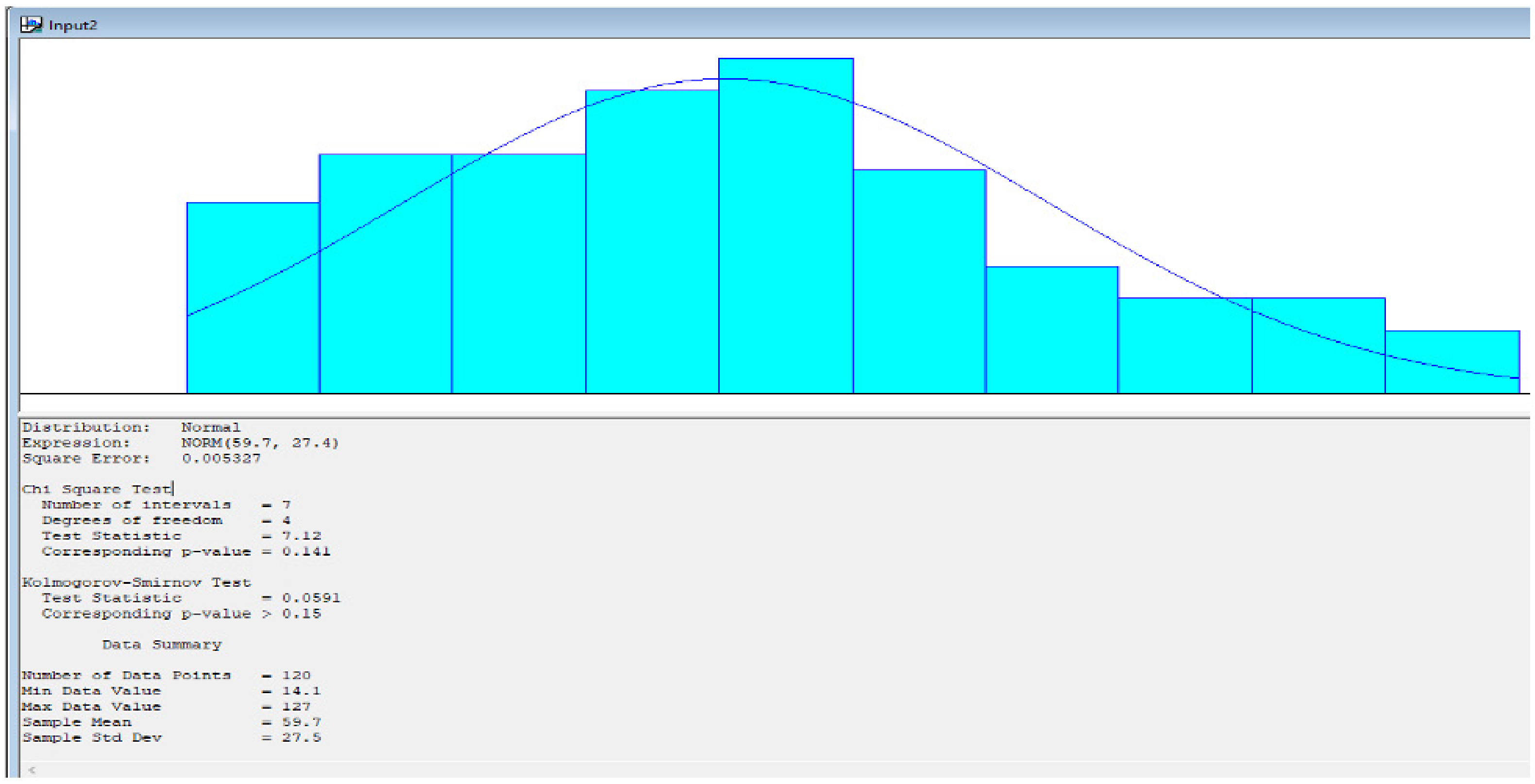


Fig. 3. Analysis of the number of fruits in a single harvest from the sample.
Source: Authors.

The coefficients of linear regression are the following:

Slope: 0.02186.
Intersect: 21.37285.

Continuing with the distribution of harvested weight, it is found that for the sample of 120 plants, the lowest recorded weight was 14.1 kg and the highest was 127 kg. The weight of harvested fruit fits a normal distribution, with a mean of 59.7 and a standard deviation of 27.4. The squared error measured was 0.005327, very small, like the parameters of the chi-square test for goodness of fit, which gives a percentile value of 0.141 and Kolmogorov-Smirnov gives a value of 0.0591. This indicates that there is no reason to reject the data as belonging to this distribution. A screenshot of said analysis made with the software input analyzer v. 10.0 can be seen in Fig. 4.



```
Distribution:    Normal
Expression:      NORM(59.7, 27.4)
Square Error:    0.005327

Chi Square Test
  Number of intervals   = 7
  Degrees of freedom    = 4
  Test Statistic        = 7.12
  Corresponding p-value = 0.141

Kolmogorov-Smirnov Test
  Test Statistic        = 0.0591
  Corresponding p-value > 0.15

        Data Summary

Number of Data Points   = 120
Min Data Value          = 14.1
Max Data Value          = 127
Sample Mean             = 59.7
Sample Std Dev          = 27.5
```

Fig. 4. Statistical analysis of the weight harvested from the sample.
Source: Authors.

C. *Simulation model*

Once the input variables were featured, the mathematical model is built, which is be based on random inputs to forecast the production of peaches, both in weight and quantity. The simulations were made under the following assumptions:

- Crop area: 1 hectare (10000 m^2).
- Plant age between 7 and 15 years.
- Thinning is done (eliminate the smallest fruits to increase nutrients to the other fruits).
- Days after defoliation: 210.

The arrangement of the plants within the crop is also a factor to take into account. Fig. 5 shows three layout schemes that were considered in the simulation. On the left side, the plants are separated from the adjacent plants by a distance of 10 m. The center image shows a rectangular layout, and the plants are separated from the adjacent ones by about 8 m. The right image shows a 6 × 8 layout, where the distance between plants measured diagonally are 8 m apart.

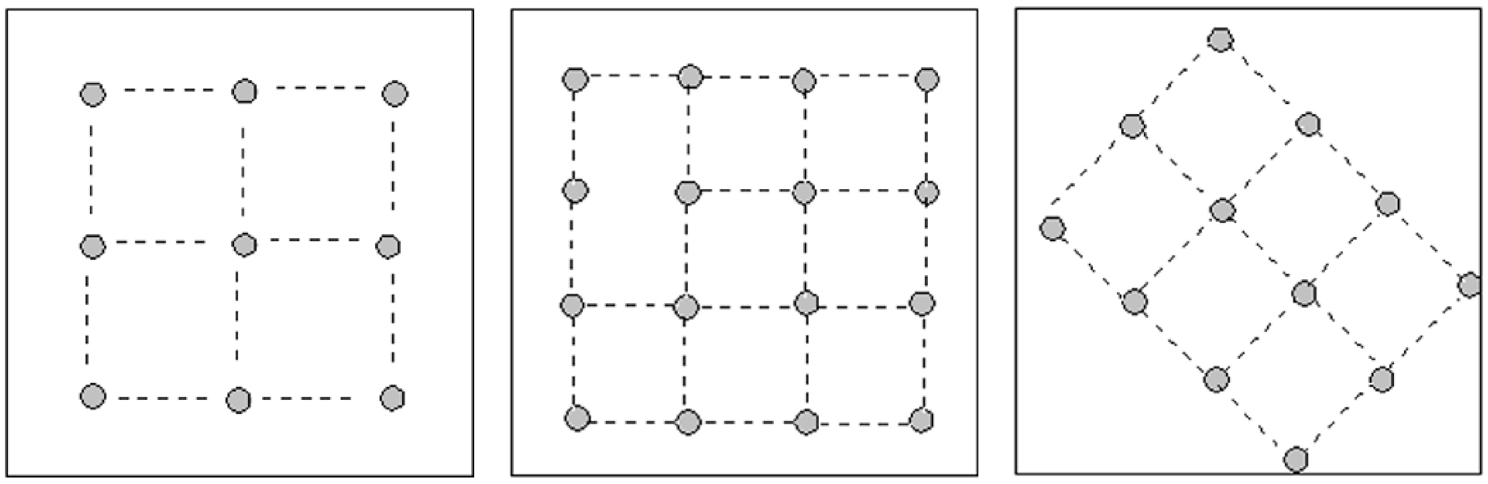


Fig. 5. Ways to arrange plants within a crop.
Source: Authors.

According to the arrangement of the plants in the crop, there will be a certain number of plants in a cultivated hectare. The numbers of plants in each arrangement per hectare were 100, 156 and 208 plants, respectively. In the model there are two fundamental output variables: the number of fruits produced by the plant, and their total weight. The distribution taken as a base is the normal one for the number of fruits and for their weight. One way to forecast the production of a farm with a known number of plants in full production is to generate a number of fruits for each plant, thus the total weight of their sum is computed. In this way, a quantity and weight of fruits would be obtained in each simulation run.

D. *Experimentation*

The assumptions for the experimentation were based on the following parameter manipulation: crop area (in hectares), crop layout (rectangular/triangular), number of plants per unit area, age of the plants and physiological characteristics (length of the branch, leaf area and equatorial diameter of the fruit).

Once these morphophysiological and agronomic parameters have been established, the software establishes the number of resulting fruits as a result, as well as the weight in kg of what is harvested. Note that the results are presented in two ways: on the one hand, there is the matrix of produced fruits (primary components of yield), and on the other hand, the matrix of production by weight of the fruit (yield). 10 replications were made for each arrangement to improve the accuracy of the model.

E. *Simulator software for agroecological production of peach fruits – Prunus. persica Jarillo variety*

Among the most important issues of the design of a software is the global conception of the system so that it can be easily developed and meet the requirements of the ecophysiological modeling of Jarillo variety peach fruit production, according to the specifications given by the model and that has been used in other crops [48]-[50]. The software was developed as a web information system, programmed in php language and with the use of forms to enter data.

F. *Hardware requirements*

The hardware required to run this simulator has the following characteristics: processor speed: greater than 2 GHz, minimum 4 GB RAM, minimum 1.6 MB disk space, and operating environment Windows 7 or higher. It is required to have a web page server installed. For php support, wampserver is needed. Once installed, it can be accessed from other machines through a local network or the Internet.

G. *Simulation results*

In order for the user to be able to modify the parameters under which he is going to carry out a simulation experiment, some input/output interfaces have been established, which work as web pages with forms which allow them to edit the parameters by evaluate as the area, crop layout, among others. The results are obtained as a production list for each plant present in the crop. For a one-hectare extension farm, these results were obtained for the 10 × 10 layout in Table 5.

TABLE 5.
SIMULATION RESULTS FOR A HECTARE AND 10 × 10 LAYOUT.

Replication	Total fruits	Average	St. deviation	Total weight	Average	St. deviation	Tons produced
1	75 699	756.992064	181.634521	5803.22036	58.0322036	9.94315247	5.80322036
2	75 042	750.420102	209.043465	5864.3463	58.643463	14.3950516	5.8643463
3	81 803	818.030446	157.277951	6093.63398	60.9363398	12.0219486	6.09363398
4	78 289	782.897813	157.523365	6269.46822	62.6946822	12.7659694	6.26946822
5	76 350	763.504375	174.983787	5946.65436	59.4665436	10.1669118	5.94665436
6	81 225	812.252434	175.852555	6195.60493	61.9560493	13.0387826	6.19560493
7	76 781	767.817332	134.485875	5927.95186	59.2795186	12.9228345	5.92795186
8	78 302	783.021684	222.209345	6054.26058	60.5426058	15.2172624	6.05426058
9	77 041	770.417458	213.894862	5969.93963	59.6993963	12.4266749	5.96993963
10	77 532	775.326343	213.984152	5898.87982	58.9887982	16.1116465	5.89887982

Source: Authors.

As can be seen for each simulation replication, the total number of fruits, the average for each plant and its standard deviation are calculated. In addition, the total weight in kg produced by the hectare, its average and standard deviation are also shown. Additionally, the yield per ton is located in each simulation. Table 6 shows the same results, applied to an 8 × 8 layout.

Table 6.
Simulation results for a hectare and 8 x 8 layout.

Replication	Total fruits	Average	St. deviation	Total weight	Average	St. deviation	Tons produced
1	119 648	766.977766	192.02847	9422.89719	58.0322036	9.94315247	9.42289719
2	122 939	788.076272	155.503259	9249.89345	59.2941888	14.2926456	9.24989345
3	119 690	767.249713	168.032907	9352.88163	59.9543694	14.8953535	9.35288163
4	123 279	790.250104	178.643508	9248.5372	59.2854949	14.4242726	9.2485372
5	121 168	776.718425	157.90587	9226.70926	59.1455722	10.1669118	9.22670926
6	123 207	789.790661	184.905003	9343.76695	59.895942	13.7047244	9.34376695
7	122 430	784.808864	160.164919	9425.21265	60.4180298	13.2015565	9.42521265
8	124 344	797.08033	157.822367	9209.24861	59.033645	14.0292399	9.20924861
9	124 282	796.683636	184.697736	9638.33198	61.7841794	12.2929623	9.63833198
10	122 639	786.151642	142.246145	9125.49006	58.4967311	12.0090674	9.12549006

Source: Authors.

As shown in the table, the production of total fruits and harvested weight is greater, due to the fact that there are more plants on the same surface. However, the averages per plant follow the same trend. Table 7 shows the results for the 6 × 8 layout.

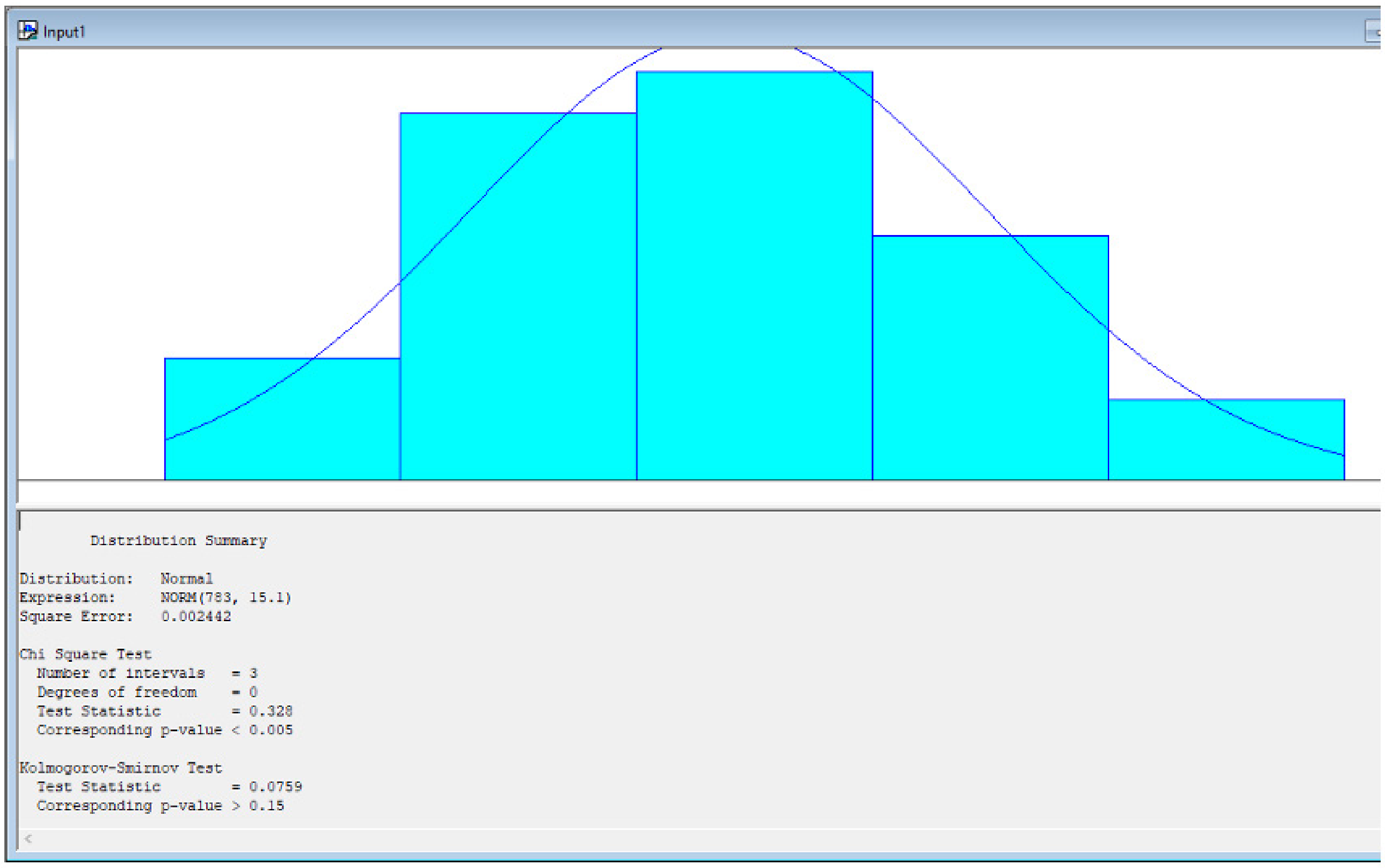
Table 7.
Simulation results for a hectare and 6 × 8 layout.

Replication	Total fruits	Average	St. deviation	Total weigth	Average	St. deviation	Tons produced
1	165 013	793.334294	179.59759	12254.6773	58.9167179	14.3934833	12.2546773
2	164 601	791.35502	152.878831	12247.0686	58.8801377	12.126021	12.2470686
3	164 691	791.787182	170.123841	12267.9307	58.9804363	11.944937	12.2679307
4	160 693	772.566048	184.393964	12397.4576	59.6031618	11.8066735	12.3974576
5	160 075	769.594152	184.142066	12885.3815	61.9489494	11.4610228	12.8853815
6	162 671	782.076736	162.721509	12551.3895	60.3432188	12.3320799	12.5513895
7	164 127	789.075038	195.044874	12318.9347	59.2256478	14.9440955	12.3189347
8	166 341	799.719887	187.62244	11934.9425	57.3795315	12.8404323	11.9349425
9	161 534	776.609927	164.350018	12391.5664	59.5748384	11.7420624	12.3915664
10	165 954	797.859671	176.15087	12587.2097	60.5154311	12.3615014	12.5872097

Source: Authors.

Given that the averages per plant are the same, and there is a high dispersion of the data in terms of average fruit and weight, the average of the averages obtained in each simulation was carried out, yielding an average per number of fruits of 782.94 and 15.35 standard deviation. For the harvested weight, about 59.69 kg per plant and 1.24 standard deviation. Classifying the data and analyzing it in input analyzer, the following behavior was obtained for the 30 simulated data of average fruit production for each plant. Fig. 6 shows this analysis.

As can be seen, the variability decreased and since the behavior of the probability distribution of production per plant is known, it is shown that it is feasible to predict production on a farm if the following variables are known: cultivated area, the arrangement of the plants, and the plant age variables within the established margins that were implemented in the simulation assumptions.



```
        Distribution Summary

Distribution:   Normal
Expression:     NORM(783, 15.1)
Square Error:   0.002442

Chi Square Test
  Number of intervals   = 3
  Degrees of freedom    = 0
  Test Statistic        = 0.328
  Corresponding p-value < 0.005

Kolmogorov-Smirnov Test
  Test Statistic        = 0.0759
  Corresponding p-value > 0.15
```

Fig. 6. Distribution of average number of fruits per plant after taking 30 simulations for 3 different hectare layouts.
Source: Authors.

IV. Conclusions

During the process of posing the problem, studying the formalisms of generating random variables and determining them, examining the samples, stochastic simulation, as well as the experimentation and analysis of results, some experiences were obtained that are stated in the following conclusions:

The performance evaluation of an agricultural process was carried out combining two formalisms: variables that affect the production of the Jarillo variety peach, and production probability distribution per plant, which makes it possible to have a mechanism for forecasting production processes. Through the simulation carried out by a software designed for this purpose, it was possible to demonstrate that it is possible to forecast the production and yield per hectare, given the phenological conditions of the crop and the arrangement of the plants inside it.

Regarding the results, they are statistically similar to the samples taken previously, which allows a yield forecast to be made based on the extension of the cultivated area, crop layout and plant age variables. With a low margin of error, the future yield of the crops can be estimated, and in this way it is easier to calculate a budget before starting the crop.

To delve deeper into this type of work it is necessary to study a better characterized sample, such as the distribution of the thermal floor, humidity conditions and variables associated with the climate, which have a more complex modeling, but which can be enriched by combining various formalisms to deal with continuous variables of the physical process, combined with probability distribution functions.

Financing

Scientific research article derived from the research project “Generation of a multivariate statistical model that allows predicting the productive behavior of the Jarillo variety peach in the province of Pamplona”, financed by “Universidad de Pamplona”. Starting year: 2015, ending year: 2018.

References

[1] FAOSTAT, “Crops and livestock products,” *FAO*. Accesed: Jun., 2021. [Online]. Available: http://www.fao.org/faostat/en/#data/QC

[2] J. Hernández, D. Micheletti, M. Bink, E. Van de Weg, C. Cantín, N. Nazzicari, A. Caprera, M. Dettori, S. Micali, E. Banchi, J. Campoy, E. Dirlewanger, P. Lambert, T. Pascal, M. Troggio, D. Bassi, L. Rossini, I. Verde, B. Quilot-Turion, F. Laurens, P. Arús & M. Aranzana, “Integrated QTL detection for key breeding traits in multiple peach progenies”, *BMC Genom*, vol. 18, no. 1, pp. 1–15, Jun. 2017. https://doi.org/10.1186/s12864-017-3783-6

[3] R, Pio, F.B.M. de Souza, L, Kalcsits, R.B. Bisi and D.D. Farias, “Advances in the production of temperate fruits in the tropics”, *Acta Sci Agron*, vol. 41, no. 1, pp. 1–10, Nov. 2019. https://doi.org/10.4025/actasciagron.v41i1.39549

[4] S. Cancino, G. Cancino y E. Quevedo, “Factores determinantes de la rentabilidad económica del cultivo de durazno en la Provincia de Pamplona, Norte de Santander, Colombia”, *Rev. Espacios*, vol. 40, no 13, pp. 1–9, Abr. 2019. Available from http://www.revistaespacios.com/a19v40n13/a19v40n13p18.pdf

[5] B. Jana, “Performance of Some Low Chill Peach, [Prunus persica (L) Batsch] Under Eastern Plateau Regions of India”, *Int J Curr Microbiol App Sci*, vol. 4, no. 12, pp. 752–757, Dec. 2015. Available from https://www.ijcmas.com/vol-4-12/B.R.Jana.pdf

[6] T. Campos, “Species y variedades de hoja caduca en Colombia”, in *Los frutales caducifolios en Colombia. Situación actual, sistemas de cultivo y plan de desarrollo*, D. Miranda, G. Fisher y C. Carranza, eds. BO, CO: Soccolhort, 2013, pp. 47–66. Disponible en http://hdl.handle.net/20.500.12324/33528

[7] *Evaluaciones Agropecuarias Municipales EVA*, MinAgricultura, Jul. 2021. https://www.datos.gov.co/Agricultura-y-Desarrollo-Rural/Evaluaciones-Agropecuarias-Municipales-EVA/2pnw-mmge

[8] S. Silva-Laya, H. Silva-Laya, y S. Pérez-Martínez, “Eficiencia energética y monetaria de sistemas de producción de durazno (Prunus pérsica) en El Jarillo, Venezuela”, *IDESIA*, vol. 35, no 4, pp. 17–26, Dic. 2017. https://doi.org/10.4067/S0718-34292017000400017

[9] S. Silva, S. Pérez & J. Álvarez, “Socioecological diagnosis and peri-urban family agriculture typification, with emphasis in the production of peach (P. persica), in El Jarillo, Venezuela”, *Rev FCA UNCuyo*, vol. 51, no 1, pp. 351–368, Jun. 2019. Available: https://revistas.uncu.edu.ar/ojs/index.php/RFCA/article/view/2456

[10] M. I. Moyano, P. Flores, S. Seta, L. Andrea y C. Severin, “Efecto de diferentes prácticas culturales sobre la producción, calidad y maduración de frutos de duraznero cv. Early Grande”, *Cienc Agron*, vol. 17, pp. 7–11, Ago. 2010. Available: https://test-cienciasagronomicas.unr.edu.ar/journal/index.php/agronom/article/view/8

[11] A. García-Mogollón y M. Torres-Zamudio, “Estudio de vigilancia tecnológica sobre el desarrollo de patentes en el campo de la producción y transformación de durazno”, *Rev Cienc Agricult*, vol. 14, no. 1, pp. 15–29, May. 2017. https://doi.org/10.19053/01228420.v14.n1.2017.6084

[12] P. Steduto, T.C. Hsiao, E. Fereres, D. Raes y FAO, *Respuesta del rendimiento de los cultivos al agua 2012*. ROM, IT: FAO, 2012. Disponible en https://www.fao.org/3/i2800s/i2800s.pdf

[13] E. J. Link, “Investigation and modeling of the optimization potential of adapted nitrogen fertilization strategies in corn cropping system with regard to minimize nitrogen losses”, *Ph. D. Dissertación*, Fac Agric Sci, UNI, ST, DE, 2005. Available from https://opus.uni-hohenheim.de/volltexte/2005/122/pdf/2005-11-30_Dissertation_JLink.pdf

[14] K. J. Boote, M. R. Rybak, J. M.S. Scholberg & J. W. Jones, “Improving the CROPGRO-Tomato Model for Predicting Growth and Yield Response to Temperature”, *HortScience*, vol. 47, no. 8, pp. 1038–1049, Aug. 2012. https://doi.org/10.21273/HORTSCI.47.8.1038

[15] J. A. Ramos, E. Becerra, J. F. Cárdenas y R. Jiménez, “Aplicación del modelo Aquacrop para un cultivo de maíz (Zea mays L)”, *Rev Sist Prod Agroecol*, vol. 10, no. 2, pp. 19–49, Nov. 2019. https://doi.org/10.22579/22484817.730

[16] M. E. Fernández, *Diagnóstico de modelos agroclimáticos evaluacion del riesgo agroclimático por sectores*. BO, CO: Fonade/IDEAM, 2013. Recuperado de http://www.ideam.gov.co/documents/21021/21138/Uso+de+Modelos+agroclim%C3%A1ticos.pdf/9f53a23d-9afa-4fda-aad3-5fe407c6cfea

[17] IICA/UE, “Elementos conceptuales básicos sobre modelos de simulación para estudiar el clima y sus impactos en la agricultur”, in *Modelos de simulación y herramientas de modelaje: elementos conceptuales y sistematización de herramientas para apoyar el análisis de impactos de la variabilidad y el cambio climático sobre las actividades agrícolas*. SJO, CR: UE/IICA, 2015, pp. 9–21. Available: http://repositorio.iica.int/handle/11324/3045

[18] C. T. De Wit, “Modelling Agricultural Production”, *Acta Hortic*, vol. 184, pp. 59–70, 1986. https://doi.org/10.17660/ActaHortic.1986.184.6

[19] B. Candelaria, O. Ruiz, F. Gallardo, P. Pérez, Á. Martínez y L. Vargas, “Aplicación de modelos de simulación en el estudio y planificación de la agricultura, una revisión”, *Trop Subtrop Agroecosystems*, vol. 14, no. 3, pp. 999–1010, Dic. 2011. Disponible en https://www.revista.ccba.uady.mx/ojs/index.php/TSA/article/view/1103

[20] A. R. Graves, P. J. Burgess, J. Palma , K. Keesman, W. van der Werf , C. Dupraz , H. van Keulen, F. Herzog & M. Mayus, “Implementation and calibration of the parameter-sparse Yield-SAFE model to predict production and land equivalent ratio in mixed tree and crop systems under two contrasting production situations in Europe”, *Ecol Modell*, vol. 221, no 13-14, pp. 1744–1756, Jul. 2010. https://doi.org/10.1016/j.ecolmodel.2010.03.008

[21] M.A. Ríos, R. Arrieta & A. Torres, “Angular Instability “Day Ahead” Risk Forecasting - Probabilistic Dependency on Load”, *IEEE LATAM Transc*, vol. 5, no. 8, pp. 585–590, Dec. 2007. https://doi.org/10.1109/T-LA.2007.4445710

[22] E. Chacón, I. Besembel y J. Hennet, *Automatización Industrial*. MDA, VE: ULA, 2001.

[23] A. Law, *Simulation Modeling and Analysis*, 5 Ed. NYC, NY, USA: McGraw-Hill, 2000.

[24] F. DiCesare & A. Desroches, “Modeling, Control and Performance Analysis of automated systems using Petri Nets”, *J Dyn Control Syst*, vol.47, Part 3, pp. 121–172, Sep. 1991. https://doi.org/10.1016/B978-0-12-012747-4.50009-3

[25] C. J. Barrera y J. C. Correa, “Distribución predictiva bayesiana para modelos de pruebas de vida vía MCMC”, *Revcoles*, vol. 31, no 2, pp. 145–155, Dic. 2008. Disponible en https://revistas.unal.edu.co/index.php/estad/article/view/29611

[26] P. H. Gutiérrez y R. S. De la Vara, *Análisis y diseño de experimentos*, 3 ed. CDMX, MX: Mc Graw Hill, 2012.

[27] IDEAM, “Metodología para la elaboración del Mapa de Ecosistemas escala 1:100.000”. Consultado el 12 de agosto de 2016. [*Online*]. Disponible en http://www.ideam.gov.co/web/ecosistemas/mapa-ecosistemas-continentales-costeros-marinos

[28] T. Lisandru, A. Füstös, V. Mitre & A. Dumitras, “Sweet Cherry (Prunus avium L.) and Peach (P. persica L.) Phenological Growth Stages According to BBCH Scale”, *Bull UASVM Hortic*, vol. 74, no 1, pp. 65–67, May. 2017. https://doi.org/10.15835/buasvmcn-hort:12361

[29] E. Fadón, M. Herrero & J. Rodrigo, “Flower development in sweet cherry framed in the BBCH scale”, *Sci Hortic*, vol. 192, pp. 141–147, Aug. 2015. https://doi.org/10.1016/j.scienta.2015.05.027

[30] O. H. Mounzer, W. Conejero, E. Nicolás, I. Abrisqueta, Y. V. García-Orellana, L. M. Tapia, J. Vera, J. M. Abrisqueta & M. del C. Ruiz-Sánchez, “Growth Pattern and Phenological Stages of Early-maturing Peach Trees Under a Mediterranean Climate”, *HortSci*, vol. 43, no. 6, pp. 1813–1818, Oct. 2008. https://doi.org/10.21273/HORTSCI.43.6.1813

[31] C. A. Schneider, W. S. Rasband & K. W. Eliceiri, “NIH Image to ImageJ: 25 years of image analysis”, *Nat Methods*, vol. 9, no. 7, pp. 671–675, Jun. 2012. https://doi.org/10.1038/nmeth.2089

[32] H. Zou & L. Xue, “A Selective Overview of Sparse Principal Component Analysis”, *Proc of the IEEE*, vol. 106, no. 8, pp. 1311–1320, Aug. 2018. https://doi.org/10.1109/JPROC.2018.2846588

[33] R. Goswami, S. Chatterjee & B. Prasad, “Farm types and their economic characterization in complex agro-ecosystems for informed extension intervention: Study from coastal West Bengal, India”, *Agric Food Econ*, vol. 2, pp. 1–24, Jul. 2014. https://doi.org/10.1186/s40100-014-0005-2

[34] J. M. Rojo, *Regresión lineal múltiple*. MAD, ES: ES, Instituto de Economía y Geografía, 2007.

[35] D. Peña, *Análisis de datos multivariantes*. MAD, ES: McGraw Hill, 2002.

[36] G. Gálvez, A. Sigarroa, T. López y J. Fernández, “Modelación de cultivos agrícolas. Algunos Ejemplos”, *Cult Trop*, vol. 31, no. 3, pp. 60–65, Jul. 2010. Disponible en https://ediciones.inca.edu.cu/index.php/ediciones/article/view/110

[37] R. H. Tirado-Malaver, J. Mendoza-Sáenz, R. Tirado-Lara y R. Tirado-Malaver, “Análisis multivariado para caracterizar y tipificar fincas productoras de papa (Solanum tuberosum L.) en Cutervo, Cajamarca, Perú”, *Trop Subtrop Agroecosyst*, vol. 24, no. 3, pp. 1–15, Jun. 2021. https://www.revista.ccba.uady.mx/ojs/index.php/TSA/article/view/3744/1677

[38] G. López, R. R. Favreau, C. Smith, & T. M. DeJong, “L-PEACH: A Computer-based Model to Understand How Peach Trees Grow”, *Horttechnology*, vol. 20, no. 6, pp. 983–990, Dec. 2010. https://doi.org/10.21273/HORTSCI.20.6.983

[39] I. Jacobson, G. Booch y J. Rumbaugh, *El proceso unificado de desarrollo de Software*. BOS, MA: Addison Wesley, 2001.

[40] L. Septar, C. Moale, I. Caplan & L. Bocioroag, “Biometric characteristics of 'Catherine Sel 1' peach cultivar in semiarid environment”, *Curr trends technol Sci*, vol. 10, no. 19, pp. 381–386, Ago. 2021. https://doi.org/10.47068/ctns.2021.v10i19.050

[41] F. Soto-Bravo y M. I. González-Lutz, “Análisis de métodos estadísticos para evaluar el desempeño de modelos de simulación en cultivos hortícolas”, *Agron Mesoam*, vol. 30, no. 2, pp. 517–534, Ago. 2019. https://doi.org/10.15517/am.v30i2.33839

[42] M. Jayakumar, M. Rajavel & U. Surendran, “Climate-based statistical regression models for crop yield forecasting of coffee in humid tropical Kerala, India”, *Int J Biometeorol*, vol. 60, no. 12, pp. 1943–1952, Dec. 2016. https://doi.org/10.1007/s00484-016-1181-4

[43] M. Scarlato, S. Dogliotti, G. Giménez, A. Borges, Ó. Bentancur y A. Lenzi, “Análisis y jerarquización de factores determinantes de las brechas de rendimiento del cultivo de frutilla en el sur del Uruguay”, *Agrocienc Urug*, vol. 21, no. 1, pp. 43–57, Jun. 2017. Disponible en http://www.ainfo.inia.uy/digital/bitstream/item/9821/1/SAD-758-p.3-9.pdf

[44] K. D. de Morais, B. Xavier, D. F. da Silva, J. A. Oliveira e C. Bruckner, “Avaliação física e química de frutos de cultivares de pessegueiro”, *Eng Agric*, vol. 25, no. 2, pp. 157–163, Abr. 2017. https://doi.org/10.13083/reveng.v25i2.712

[45] J. P. Cremasco, R. G. Matias, D. F. da Silva, J. A. Oliveira e C. Bruckner, “Qualidade pós-colheita de oito variedades de pêssego”, *Comun Sci*, vol. 7, no. 3, pp. 334–342, Dez. 2016. https://doi.org/10.14295/cs.v7i3.1404

[46] R. G. Matias, D. F. da Silva, J. O. e Silva, J. A. Oliveira, J. P. Cremasco e C. Bruckner, “Qualidade de nectarinas produzidas em região de clima subtropical”, *Rev Ceres*, vol. 62, no. 6, pp. 621–626, Dez. 2015. https://doi.org/10.1590/0034-737X201562060016

[47] R. G. Matias, C. Bruckner, D. F. da Silva, P. C. Carneiro & J. A. de Oliveira, “Adaptability and stability of peach and nectarine cultivars in subtropical climate”, *Rev Ceres*, vol. 64, no. 5, pp. 516–522, Oct. 2017. https://doi.org/10.1590/0034-737X201764050009

[48] A. Lira-Conde, E. Ocaranza-Sánchez, J. Cadena-Iñiguez, L. Tapia-López, M. M. Solís-Oba y A. Ruiz-Font, “Aplicación de un modelo matemático para predecir el rendimiento vegetal de un compuesto de interés industrial”, *Agroproduct*, vol.11, no. 9, pp. 59–67, Sep. 2018. https://doi.org/10.32854/agrop.v11i9.1216

[49] J. A. S. Escalante-Estrada, , M. T. Rodríguez-González y Y. I. Escalante-Estrada, “Modelos que describen la distribución del rendimiento, sus componentes y radiación solar en ayocote en espaldera de tripie”, in *Ciencias Matemáticas aplicadas a la Agricultura*., F. Pérez-Soto, D. Sepúlveda-Jiménez, R. Salazar-Moreno y D. Sepúlveda-Robles, eds, CDMX, MX: ECORFAN, 2017, pp. 9–17. Recuperado de https://www.ecorfan.org/handbooks/Ciencias%20Matematicas%20aplicadas%20a%20la%20Agronomia%20T-I/HCMA_TI.pdf

[50] A. C. Gomes, A. Robaina, M. Peiter, F. Soares e A. R. Parizi, “Modelo para estimativa da produtividade para a cultura da soja”, *Ciênc Rural*, vol. 44, no. 1, pp. 43–49, Jan. 2014. https://doi.org/10.1590/S0103-84782013005000145

[51] *SPSS Statistics*. (v. 25), IBM. [Online]. Available: https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-25

[52] *R Project for Statistical Computing*. (v. 42.1), R Development Core Team. [Online]. Available: https://www.r-project.org/

[53] *Arena Input Analyzer software*. (v. 16.0), Rockwell Automation. [Online]. Available: https://www.rockwellautomation.com/en-us/products/software/arena-simulation.html

Carlos Arturo Parra Ortega is a Systems Engineer, graduated from the Industrial University of Santander (Colombia). Master’s degree in Computer Science. Doctor of Applied Sciences from the Universidad de Los Andes (Venezuela). Teacher ranked as Full Professor at the University of Pamplona and member of the Research Group in Computational Sciences (CICOM), attached to the Faculty of Engineering and Architecture. His research area is related to modeling and simulation, industrial automation and multi-agent systems. https://orcid.org/0000-0003-3593-9504

Enrique Quevedo García is an Agricultural Engineer, graduated from the National University (Colombia). Master’s degree in Agricultural Sciences from the National University (Colombia). Doctor in Natural Sciences and Development with an emphasis on Agricultural Production Systems at the National University of Costa Rica. He is classified as a Research Associate in Minciencias. He worked as a type 2 professional at the Colombian Corporation for the Amazon, between 1989 and 1994. Also at Agrosavia, at the El Mira research center in Tumaco from 1997 to 2001 as an associate researcher. At the University of Cundinamarca in 1996 and 1997, and from 2001 to 2003. At Uniagraria from 2001 to 2003. He is currently Professor of the Agricultural Engineering program at the Faculty of Agricultural Sciences from 2004 to the present day, at the University of Pamplona. He is part of the Research Group on Livestock and Sustainable Agriculture GIAS and works in the line of research in Crop Physiology and Plant Health. https://orcid.org/0000-0001-9465-8126

Cesar Villamizar Quiñones is an Agricultural Engineer, graduated from UPTC (Colombia). Master’s degree in Agricultural Sciences from the University of Chile (Chile). He directed the Research Center for the Panela Agroindustry CIMPA of Agrosavia, between 1995 and 2004. He is currently an associated professor of the Agronomic Engineering program, of the Faculty of Agricultural Sciences, of the University of Pamplona. He is part of the Research Group on Livestock and Sustainable Agriculture GIAS and works in the lines of research on Production Systems and Technological Information and Extension and Rural Development. https://orcid.org/0000-0001-5937-5278