

# NEURONAL NETWORKS TO DESCRIBE EPIDEMICS OF COCOA WITCHES' BROOM

## Edson Ampélio Pozza

Federal University of Lavras (UFLA), Phytopatology Dept., CP 3037, CEP 37.200-000, Lavras, MG, Brazil  
Corresponding author: eapozza@ufla.br

## Luiz Antônio Maffia

Federal University of Viçosa (UFV), Phytopatology Dept., Campus Universitário, S/N, CEP 36.570-000, Viçosa, MG, Brazil

## Carlos Arthur Barbosa da Silva

Federal University of Viçosa (UFV), Food Technology Dept., Campus Universitário, S/N, CEP 36.570-000, Viçosa, MG, Brazil

## Marcelo de Carvalho Alves

Federal University of Lavras (UFLA), Engineering Dept., CP 3037, CEP 37.200-000, Lavras, MG, Brazil

## José Luis Braga

Federal University of Viçosa (UFV), Informatics Dept., Campus Universitário, CEP 36.570-900, Viçosa, MG, Brazil

## João de Cássia do Bonfim Costa

CEPEC/CEPLAC/MAPA, Phytopathology Section, Rua G, Setor Sudoeste, Campus do Inmet, CEP 70.680-900, Cruzeiro, DF, Brazil

**Abstract:** Artificial Neural networks (ANN) were evaluated as tools to describe epidemics of cocoa witches' broom and as a potential method to forecast the disease. The ANN were built with data collected in Altamira-PA-Brazil, between January 1986 and December 1987, and were compared by regression analysis. The variables studied were basidiocarp production, disease intensity, and 16 climatic variables. Seven climatic variables were selected at 1 to 10 weeks before basidiocarp production and 11 variables at the 8th and 9th weeks before evaluation of disease intensity. Temporal series were also analyzed. A total of 37 regression models were tested and 100 ANN built. Neuronal networks could forecast disease intensity more efficiently than regression equations. The best ANN used 11 climatic variables, in the 9th week before disease occurrence. The best ANN, with two intermediary layers of artificial neurons, and regression equation to describe basidiocarp production included the variable rainfall duration, in hours.

**Index terms:** Soft computing, epidemiology, forecast, plant disease.

Received: May 24, 2018 - Accepted: June 25, 2018

## INTRODUCTION

Cocoa (*Theobroma cacao* L.) is originated from the rain forests of Tropical America, where it is still found in wild state. In some countries of Latin America and Asia, it constitutes one of the main economic supports for the population as well as for export FAO (2015). The witches' broom disease, caused by the fungus *Moniliophthora perniciosa* (Stahel) Aime & Phillips-Mora, threatens the cocoa production. Witches' broom, considered the most destructive disease of cacao farms Bastos (1991), is spread throughout all cocoa producing countries of South America,

in the islands of the Caribbean and in Central America. In the Brazilian Amazon, it causes losses of up to 70% in the production of commercial plantations older than six years (Bastos, 1990). In Bahia, the largest producing area in Brazil, the occurrence of the disease is knocking down the farms and consequently the Atlantic rain forest, what leads to serious ecological damages, unemployment and export reduction (Costa et al. 2010). In this environment, the climatic variables are important to explain the disease progress curve of the witches' broom.

Artificial Neuronal Networks (ANN) are an alternative to analyze data, which is not adapted

to the statistical methods and chaotic systems (Eberhart and Dobbins, 1990; Liu, Wu and Huang, 2010). The ANN are based in studies of the brain and the nervous system and their basic objective is "to learn" from its own experience (Rumelhart, Widrow and Lehr, 1994; Hertz et al. 2018). Arizmendi et al. (1993) used ANN to forecast occurrence of pollen grains in the air, which in high levels can cause epidemics of allergy. Their ANN successfully foresaw the day of the year with high concentration of pollen grain. Forecast of temporal series is one of the main applications of NN in several knowledge levels (Arizmendi et al. 1993; Silva and Silva, 1995). Collins (1993a) (1993b) mention some uses of the NN in biology and biotechnology, mainly in the patterns recognition and forecast. The recognition of patterns in a series of data (numbers in a database, spectral information or photographic images) has crucial importance for biology and agriculture. Nater, Nater and Baker (1992) accomplished studies to reproduce images of roots in soil and obtained a good performance of the system to recognize image characteristics. Das and Evans (1992) tested the ability of mechanical eyes (Machine Vision II) to detect dead embryos or unfertile eggs destined to incubation. They found 93.5% and 93.9% precision in the classification during the third and fourth days of incubation, respectively. In plant diseases, many researches describe the abilities of ANN to recognize patterns (Alves et al. 2017; Goodridge et al. 2017).

Pinto et al. (2002) describe the epidemics of coffee rust in Brazil with an ANN and Paul and Munkvold (2016) obtained a good ANN performance for the prediction of gray leaf spot epidemics in maize. Therefore, ANN was evaluated in this paper as a tool to describe epidemics and as a potential forecaster of disease occurrence. The cocoa witches' broom was used as a model of disease.

## MATERIAL AND METHODS

### Data collection

Experimental data of Costa (1993), collected in a commercial cocoa farm in Altamira-PA-Brazil, were used.

### Production of basidiocarp of *Crinipellis pernicioso* and intensity of witches' broom disease in vegetative organs of the cocoa tree

Seven hundred witches' brooms were placed in the field in 1985 and 1986 and the following variables were evaluated weekly:

TB86 = total of basidiocarps counted in 1986, in brooms placed in 1985;

TB86a = total of basidiocarps counted in 1987, in brooms placed in 1985;

TB87 = total of basidiocarps counted in 1987, in brooms placed in 1985 and 1986.

In vegetative organs of the cocoa trees, the following variables were evaluated at every 15 days:

AB = total of axillary brooms;

TB = total of terminal brooms;

BSC = total of branches with swellings and cankers;

PPS = total of petioles and pulvinus with swellings.

IWB = intensity of witches' broom (AB + TB + BSC + PPS).

### Meteorological data

From a weather station placed in the experiments, the following variables were evaluated weekly:

Tmin = average of the minimum temperature, in °C;

Tmax = average of the maximum temperature, in °C;

T25 = number of hours with temperature  $\geq 25^{\circ}\text{C}$ ;

T30 = number of hours with temperature  $\geq 30^{\circ}\text{C}$ ;

URmin = average of the minimum relative humidity, in %;

URmax = average of the maximum relative humidity, in %;

UR70 = number of hours with relative humidity  $\leq 70\%$ ;

UR95 = number of hours with relative humidity  $\geq 95\%$ ;

TR = total of rainfall, in mm;

DR = duration of the rainfall, in hours;

NDR = number of days with rainfall;

NDR2 = number of days with rainfall  $\leq 2$  mm;

NDR25 = number of days with rainfall  $\geq 2$  and  $\leq 5$  mm;

NDR58 = number of days with rainfall  $\geq 5$  and  $\leq 8$  mm;

NDR811 = number of days with rainfall  $\geq 8$  and  $\leq 11$  mm;

NDR11 = number of days with rainfall  $\geq 11$  mm.

## Data Analysis

### Selection of climatic variables related to basidiocarp production and regression analysis

According to the coefficient of determination ( $R^2$ ) and the highest frequency of significant coefficients of partial determination, seven meteorological variables among the 16 studied by Costa (1993), were selected: NDR, DR, NDR2, NDR58, NDR11, NDR811, NDR25, during the intervals of 1 to 10 weeks before disease occurrence. These intervals were chosen because they contained the largest number of the explanatory climatic variables, according to Costa (1993).

To evaluate the influence of the seven climatic variables selected on total basidiocarp production, the following regression model was used:

$$Y = b_0 + b_1WK1 + b_2WK2 + \dots + b_{11}WK10 \quad (1)$$

The dependent variable  $Y$  is the total basidiocarp production in brooms in 1986 (TB86a) and in 1987 (TB87);  $b_0$  is the regression constant;  $b_1, b_2, \dots, b_{10}$  are the coefficients of partial regression, and WK1, WK2, ..., WK10 are the values of the climatic variables at the weeks 1 to 11 before assessment of basidiocarp production.

Other model was also evaluated in which the seven climatic variables chosen were considered the independent variables for a certain week:

$$Y = b_0 + b_1DR + \dots + b_7NDR25 \quad (2)$$

Thus, for the first, second, 52nd weeks before evaluating basidiocarp production, the independent variables corresponded to NDR, DR, ..., NDR25.

### Selection of climatic variables related to witches' broom intensity and regression analysis

Eleven climatic variables from the 16 studied by Costa (1993) were selected:  $T_{min}$ ,  $T_{max}$ ,  $UR_{min}$ , T25, T30, UR95, UR70, NDR 25, NDR 811, NDR 11 and NDR. They were selected because they were significant in the regression equations developed by Costa (1993). The interval between 8 and 9 weeks before disease assessment was chosen because of the largest frequency of significant associations between the climatic variables and disease intensity in a multivariate analysis Costa (1993). A model of multiple regression with the variables chosen was used:

$$Y = b_0 + b_1T25 + \dots + b_{11}NDR \quad (3)$$

The dependent variable  $Y$  is intensity of witches' broom (IWB);  $b_0$  is the constant of regression;  $b_1, b_2, \dots, b_{11}$  are coefficients of partial regression and T25, T30, ..., NDR are the independent variables described above.

In an attempt to select a better adjusted model, the observations of weeks 8 and 9 were grouped in a new model:

$$Y = b_0 + b_1T25 + \dots + b_{11}NDR + b_{12}WK \quad (4)$$

The independent binary variable WK represents the observation in week 8 (WK=0) or week 9 (WK=1).

The significance of the coefficients of partial regression was evaluated by the t test ( $P=0.05$  and  $P=0.01$ ) and those non-significant were eliminated. The coefficient of determination, error mean square and the significance of F test ( $P=0.05$  and  $P=0.01$ ) selected the best regression equations.

### Construction of neuronal networks

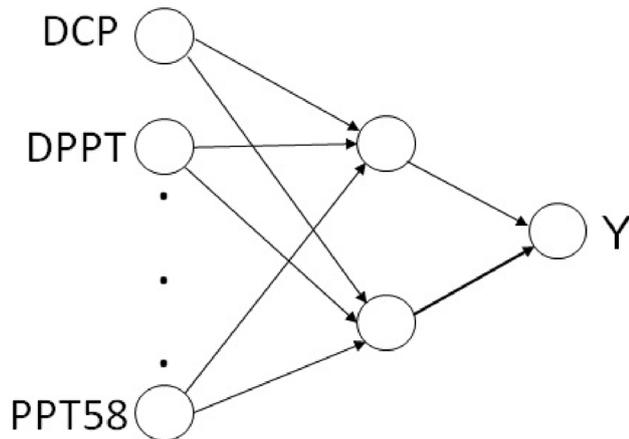
The data were organized and the ANN were constructed and analyzed with the software BRAINCEL 3.0 (Promised Land Technologies, Inc.).

**Neuronal networks used to describe basidiocarp production by *Crinipellis perniciosa***

Data of basidiocarp production from January 1986 to December 1987, with 96 observations, were used.

The construction of ANN requires the definition of a topology, with the elements of the input layers, intermediary layers, and output. Two models were built in terms of the input layer, whereas for both the output layer was basidiocarp production. The program it self determined the number of artificial neurons and intermediary layers.

The first model considered the climatic variables delayed by weeks before basidiocarp production. The input layer corresponded to the independent variables NDR, DR,... , NDR 58, where as the output layer would, correspond to basidiocarp production in brooms placed and counted in 1986 and counted in 1987 (Figure 1).

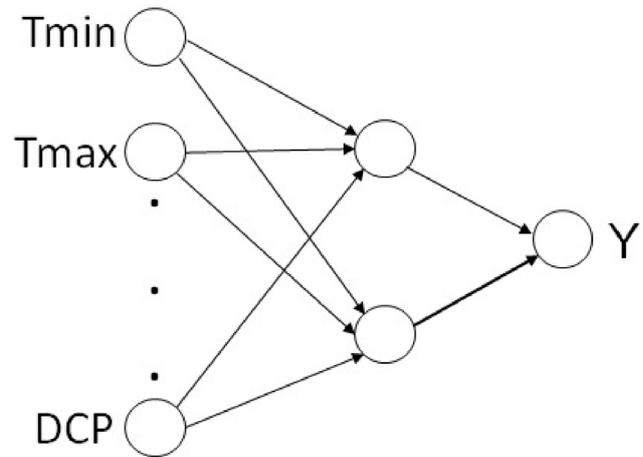


**Figure 1:** Representation of a neural network to describe the production of basidiocarps in the first week. The independent variables DCP, DPPT PPT25 were used to represent the input layer and, the production of basidiocarps (Y), the output layer.

The second model considered weeks delayed by variables before basidiocarp production. The input layer considered the values of one variable from the first to the 11<sup>th</sup> week (Figure 2).

A temporal series for basidiocarp production was also used to construct the ANN. From literature reports Arizmendi et al. (1993); Silva and Silva (1995), a recursive model was used. In

this model, the basidiocarp production in certain period would be explained by the previous first, second, third or fourth weeks. Therefore, more four nets were built.



**Figure 2:** Representation of a neural network to describe the intensity of the witches' broom in the 8<sup>th</sup> week, The variables Tmin, Tmax, DCP were used to represent the input layer and, the intensity of the disease, the output layer.

Since the highest determination coefficients in the regression analysis were for the variable DR, ANN were built where the input layer consisted simultaneously of observations of DR and other climatic variables. An artificial neuron was included in the input layer with a binary representation of the variable: DR's value was 1 and the other variables was 0.

The specification of a subset of the data is requested for training of the network and another subset is requested for tests. About 75% of the original observations were randomly selected for training, while 25% for testing it. The tolerance (fraction of the interval between the highest and the lowest values of the output layer data) of 0.05 was considered to establish whether a value calculated by the net was acceptable.

**Neuronal networks to describe witches' broom intensity**

To construct ANN for witches' broom intensity, the independent variables used in the input layer were selected for regression analysis (Tmin, Tmax,... , NDR) and disease intensity was the output layer. The networks were evaluated

with one and two intermediary layers. The other characteristics were identical to the networks built for basidiocarp production.

### Comparison between linear regression and neuronal networks

The mean error of forecast (MEF) and the mean square error (MSE) were used to compare regression to the ANN where:

$$MEF = [S(X_{obs} - X_{calc}) * 100 / (X_{obs} + 1)] / n \quad (5)$$

$$MSE = [S(X_{obs} - X_{calc})^2] / n \quad (6)$$

where  $\Sigma$  is the i sum to n,  $X_{obs}$  the value of basidiocarp production or disease intensity observed,  $X_{calc}$  the value of the regression or the net neuronal calculated, and n the total of observations.

## RESULTS AND DISCUSSION

### Regression analysis for basidiocarp production and witches' broom intensity

For basidiocarp production, the regression models with highest determination coefficients

considered seven weather variables always including the variables DR and NDR58. A larger number of significant interactions took place in the first and the seventh weeks before disease occurrence (Table 1).

For disease intensity, all equations that included the weeks from 1 to 10 before disease occurrence  $R^2$  lower than 0.45. The F test was not significant for the weeks 3, 9, and 10 before disease occurrence. The variables DR and NDR11 showed larger number of significant interactions, 5 and 4 respectively (Table 2).

For all regression models that included 11 weather variables and witches' broom intensity, the values of  $R^2$  were lower than 0.57, although significant for the t test ( $P = 0.01$ ). The variables T25 and NDR11 were significantly included with higher frequency, although T25 was not included in the equation for weeks 8 and 9. The other significant variables were Tmax and T30 ( $P=0.01$ ), in the equation 89 (Table 3).

### Neuronal networks

The ANN that considered the weather variables one-week before basidiocarp production, were built with two intermediary layers. Lowest MSE and MEF were for the variables DR and NDR2. For the ANN with an intermediary layer for NDR2, the MEF was 28.58%, although MSE was high.

**Table 1:** Coefficients of determination ( $R^2$ ) and partial regression between seven climatic variables per week and the production of basidiocarps of the witches' broom in Altamira, Pará, Brazil.

Climatic variable	Time interval in weeks, after the observations of climate variables to the production of basidiocarps										Model	
	1	2	3	4	5	6	7	8	9	10	$R^2$	adjusted $R^2$
DCP	17.30**	-2.38 <sup>ns</sup>	3.36 <sup>ns</sup>	16.14 <sup>ns</sup>	4.46 <sup>ns</sup>	10.93 <sup>ns</sup>	16.53*	2.44 <sup>ns</sup>	-0.88 <sup>ns</sup>	-10.04 <sup>ns</sup>	0.52**	0.44**
DPPT	-16.38 <sup>ns</sup>	-9.35**	-7.75*	-2.52 <sup>ns</sup>	1.97 <sup>ns</sup>	8.19**	14.94**	10.57**	9.19**	-0.28 <sup>ns</sup>	0.70**	0.69**
PPT11	34.86**	1.05 <sup>ns</sup>	1.75 <sup>ns</sup>	-8.76 <sup>ns</sup>	-15.97 <sup>ns</sup>	8.68 <sup>ns</sup>	24.44*	19.83 <sup>ns</sup>	18.66 <sup>ns</sup>	6.9 <sup>ns</sup>	0.48**	0.36**
PPT2	-23.09**	6.46 <sup>ns</sup>	-7.66 <sup>ns</sup>	-8.15 <sup>ns</sup>	0.17 <sup>ns</sup>	-16.81*	-19.3**	-1.81 <sup>ns</sup>	-1.87 <sup>ns</sup>	14.11**	0.56**	0.52**
PPT811	61.6 <sup>ns</sup>	47.18 <sup>ns</sup>	36.47 <sup>ns</sup>	32.93 <sup>ns</sup>	-4.14 <sup>ns</sup>	-22.35 <sup>ns</sup>	20.21 <sup>ns</sup>	-15.11 <sup>ns</sup>	-13.22 <sup>ns</sup>	-3.75 <sup>ns</sup>	0.13 <sup>ns</sup>	—
PPT25	49.39**	18.34 <sup>ns</sup>	43.39*	36.17 <sup>ns</sup>	40.44**	57.39**	17.07 <sup>ns</sup>	6.75 <sup>ns</sup>	5.96 <sup>ns</sup>	-30.13 <sup>ns</sup>	0.50**	0.41**
PPT58	49.29**	21.13 <sup>ns</sup>	14.29 <sup>ns</sup>	35.39*	33.23*	56.53**	70.09**	60.63**	10.08 <sup>ns</sup>	-4.44 <sup>ns</sup>	0.64**	0.62**

ns, \* and \*\* - not-significant and significant at the levels of 0.05 and 0.01 probability, respectively, by t test (partial regression coefficient) and F test ( $R^2$  and adjusted  $R^2$ ), DCP = number of days with precipitation; DPPT = rainfall duration in hours; PPT11 = number of days with rainfall > 11 mm; PPT2 = number of days with rainfall < 2 mm; PPT811 = number of days with rainfall > 8 and < 11 mm; PPT25 = number of days with rainfall > 2 and > 5 mm and PPT 58 = number of days with rainfall > 8 and < 11 mm.

**Table 2:** Coefficients of determination ( $R^2$ ) and partial regression from 1 to 10 weeks out of phase by climatic variable and the production of basidiocarps of the witches' broom in Altamira, Pará, Brazil.

Weeks before basidiocarp production	Climatic variable							Model	
	dppt	ppt2	ppt25	PPT58	PPT811	PPT11	DCP	Complete	Reduced
	Partial regression coefficients							$R^2$	adjusted $R^2$
1	14.79**	32.26 <sup>ns</sup>	53.39 <sup>ns</sup>	63.21 <sup>ns</sup>	49.89 <sup>ns</sup>	100.96**	20.91 <sup>ns</sup>	0.42*	0.30**
2	-20.41**	33.31 <sup>ns</sup>	45.29 <sup>ns</sup>	31.36 <sup>ns</sup>	37.26 <sup>s</sup>	88.91*	29.22 <sup>ns</sup>	0.32**	0.17**
3	-7.22 <sup>ns</sup>	15.53 <sup>ns</sup>	45.55 <sup>ns</sup>	-3.51 <sup>ns</sup>	14.19 <sup>ns</sup>	36.13 <sup>ns</sup>	23.65 <sup>ns</sup>	—	—
4	8.47 <sup>ns</sup>	-2.54 <sup>ns</sup>	5.13 <sup>ns</sup>	-19.31 <sup>ns</sup>	-39.11 <sup>ns</sup>	-54.70 <sup>ns</sup>	38.73*	0.31*	0.21 <sup>ns</sup>
5	10.96*	-27.16 <sup>ns</sup>	-18.82 <sup>ns</sup>	-34.51 <sup>ns</sup>	-66.24 <sup>ns</sup>	-100.42*	40.07*	0.35**	0.31**
6	11.82*	-76.39*	-65.55 <sup>ns</sup>	-77.27 <sup>ns</sup>	-141.99*	-115.68*	31.43*	0.45**	0.42**
7	10.99*	-22.07 <sup>ns</sup>	-46.75 <sup>ns</sup>	-2.96 <sup>ns</sup>	-45.99 <sup>ns</sup>	-46.98 <sup>ns</sup>	26.11 <sup>ns</sup>	0.42**	0.33**
8	92.29 <sup>ns</sup>	-36.44 <sup>ns</sup>	-79.74 <sup>ns</sup>	-12.56 <sup>ns</sup>	-100.25*	-65.47 <sup>ns</sup>	27.76 <sup>ns</sup>	0.33**	0.19 <sup>ns</sup>
9	9.09 <sup>ns</sup>	9.89 <sup>ns</sup>	-44.56 <sup>ns</sup>	-22.66 <sup>ns</sup>	-39.22 <sup>ns</sup>	-16.19 <sup>ns</sup>	30.73 <sup>ns</sup>	—	—
10	7.58 <sup>ns</sup>	10.84 <sup>ns</sup>	-32.47 <sup>ns</sup>	-17.45 <sup>ns</sup>	14.29 <sup>ns</sup>	-23.40 <sup>ns</sup>	34.57 <sup>ns</sup>	—	—

ns, \* and \*\* - not-significant and significant at the levels of 0.05 and 0.01 probability, respectively, by t test (partial regression coefficient) and F test ( $R^2$  and adjusted  $R^2$ ), DCP = number of days with precipitation; DPPT = rainfall duration in hours; PPT11 = number of days with rainfall > 11 mm; PPT2 = number of days with rainfall < 2 mm; PPT811 = number of days with rainfall > 8 and < 11 mm; PPT25 = number of days with rainfall > 2 and > 5 mm and PPT 58 = number of days with rainfall > 8 and < 11 mm.

**Table 3:** Coefficients of determination ( $R^2$ ) and partial regression coefficients between the 11 weeks prior to the observation of the intensity of the witches' broom and climate in Altamira, Pará, Brazil.

Time intervals in weeks	Climatic variables										Models		
	Tmin	Tmax	URmin	T25	T30	UR95	UR70	PPT25	PPT811	PPT11	DCP	Complete	Reduced
	Partial regression coefficients										$R^2$	adjusted $R^2$	
8	-0.13 <sup>ns</sup>	-1.85 <sup>ns</sup>	0.05 <sup>ns</sup>	0.12**	0.12 <sup>ns</sup>	-0.04 <sup>ns</sup>	0.02 <sup>ns</sup>	0.68 <sup>ns</sup>	-0.65 <sup>ns</sup>	0.90*	0.34 <sup>ns</sup>	0.51**	0.22**
9	-0.03 <sup>ns</sup>	-1.66 <sup>ns</sup>	-0.07 <sup>ns</sup>	0.12**	0.10 <sup>ns</sup>	-0.05 <sup>ns</sup>	-0.09 <sup>ns</sup>	0.76 <sup>ns</sup>	-0.59 <sup>ns</sup>	0.91*	0.31 <sup>ns</sup>	0.57**	0.32**
8+9	-0.08 <sup>ns</sup>	-1.74*	-0.06 <sup>ns</sup>	-	0.10*	-0.04 <sup>ns</sup>	0.58 <sup>ns</sup>	0.72 <sup>ns</sup>	-0.63 <sup>ns</sup>	0.90*	0.32 <sup>ns</sup>	0.51**	0.31*

ns, \* and \*\* - not-significant and significant at the levels of 0.05 and 0.01 probability, respectively, by t test (partial regression coefficient) and F test ( $R^2$  and adjusted  $R^2$ ), Tmin = average minimum temperature; Tmax = average maximum temperature (°C); URmin = average minimum temperature; T25 = number of hours with temperature ≥ 25°C; T30 = number of hours with temperature ≥ 30°C; UR95 = number of hours with relative humidity ≥ 95%; UR70 = number of hours with relative humidity ≤ 70%; PPT25 = number of days with rainfall ≥ 2 and ≤ 5 mm; PPT811 = number of days with rainfall ≥ 8 e ≤ 11 mm; PPT11 = number of days with rainfall ≥ 11 mm; DCP = number of days with rainfall.

The ANN used to describe basidiocarp production in the weeks 1 to 10 before disease occurrence did not provide satisfactory results. Values of MSE and MEF were high for all the ANN tested. The lowest MEF was 52.43% for the data of weeks 5 and 6 in the new inoculum sources of 1986 and old of 1987 with a MSE larger than 20.000.

The temporal series did not provide satisfactory results. Lowest MEF was 18.97%,

for one and seven weeks before basidiocarp production, for the old inoculum sources. The use of the variable DR plus the others as inputs did not present good results either. MSE and MEF were high for six ANN tested.

For three ANN built for disease intensity, MSE's were low and MEF was 3.01 and 9.61 for the week 9 and weeks 8+9, respectively. Lowest MEF was obtained at week 9, for both disease intensity and basidiocarp production.

### Comparison between regression analysis and neuronal networks to describe basidiocarp production and intensity of witches' broom

Artificial Neuronal networks and regression models obtained lower MSE and MEF for the intensity of the disease (Table 4). High adjustment level was in week 9 (Figure 3).

In the regression analysis for basidiocarp production, lower MEF (11.98%) was with NDR (Figure 4).

For the ANN, lowest MEF was obtained with DR (12.56%) with the topology 10-5-2-1 (Table 5).

In the regression analysis for the weeks delayed for variable, both MEF and MSE were too high (Table 6).

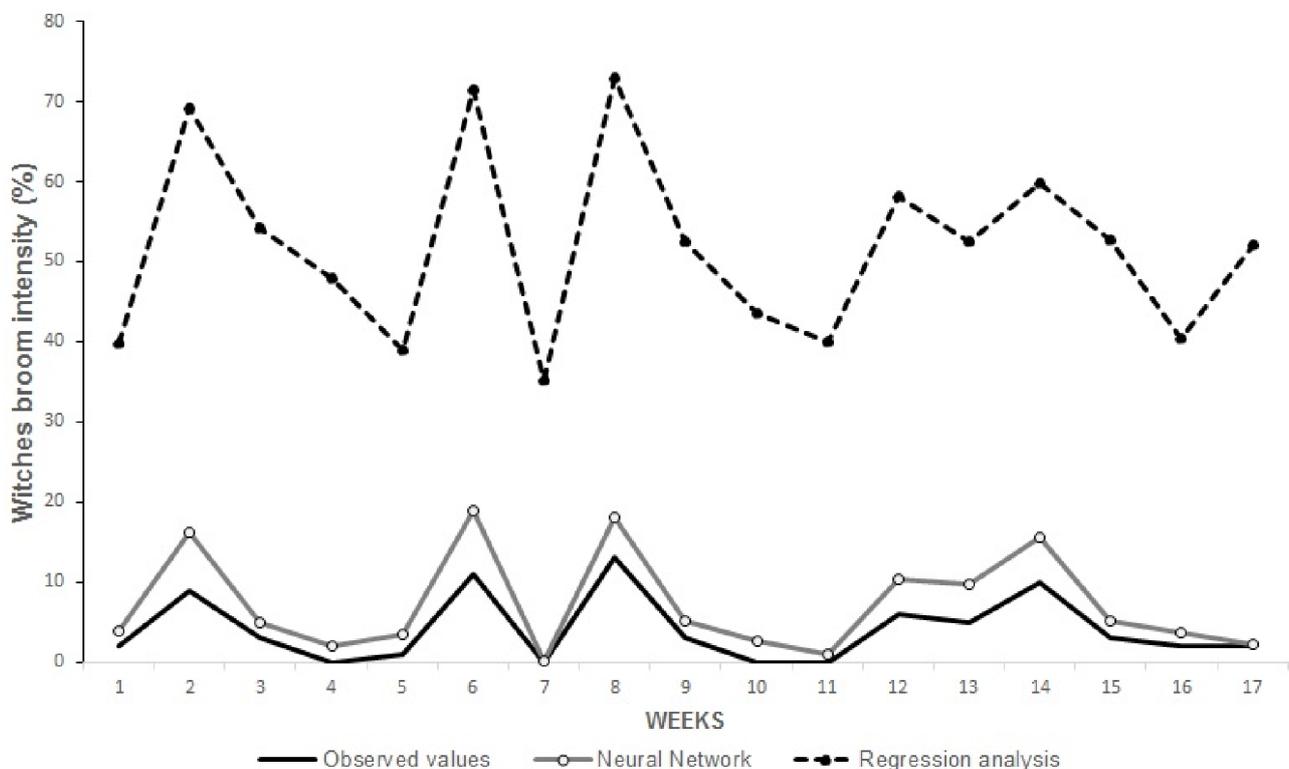
Good adjustment for basidiocarp production was found when using DR variable (Figure 5).

The most descriptive variables were DR (duration of rainfall, in hours) and NDR58 (number of days with rainfall between 5 and 8 mm) in the regression analysis for basidiocarp production. It is already known that rainfall is the most important meteorological factor for basidiocarp production (Bastos and Silva, 1980; Aranzazu, 1990). Costa (1993) found a high

**Table 4:** Intensity of the witches' broom of cocoa: comparison between regression analysis and neural networks.

Time intervals in weeks	Regression				Neural network	
	Model		Root mean square error	Prediction mean error (%)	Root mean square error	Prediction mean error (%)
	Complete	Reduced				
	R <sup>2</sup>	adjusted R <sup>2</sup>				
8	0.51**	0.22**	114.56	82.70	4.735	38.37
9	0.57**	0.32**	97.32	1549.40	32.06	3.01
89	0.51**	0.31*	86.15	16.80	12.27	9.61

\* and \*\* - significant between the levels 1 and 5%, respectively, by the F test (R<sup>2</sup> and adjusted R<sup>2</sup>).

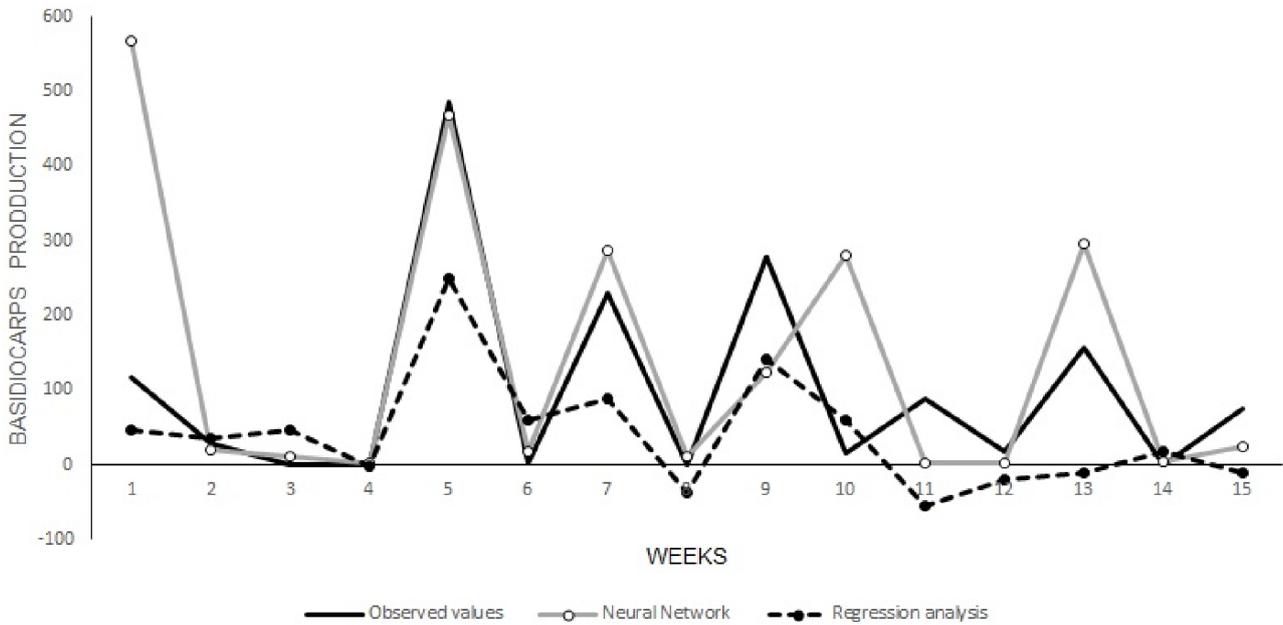


**Figure 3:** Observed values of witches' broom intensity versus predicted values by neural networks and regression model.

coefficient of determination (0.86) for DR during twenty weeks before basidiocarp production. The variable NDR (days with rainfall) presented the lowest MEF (11.98%) in all the adjusted equations. Lowest MSE (12.56%) was found in the ANN that used DR as input with the topology 10-5-2-1.

There are some alternatives to reduce the MEF of the ANN, such as increasing the number of intermediary layers, artificial neurons in the

intermediary layers, and new variables, to best represent the phenomenon. The selection of new variables related to weather, pathogen, or host is one of the main phases in the construction of the ANN. The main problem is to choose the variables that best represent a phenomenon, not excluding the importance of determining the number of intermediary ANN layers (Pinto, 2002). The human brain automatically discards the unnecessary variables to obtain an answer



**Figure 4:** Observed values of basidiocarps production versus predicted values by neural networks and regression model.

**Table 5:** Basidiocarps production of the witches' broom of chocolate tree: comparison between regression analysis and neural networks for the week delayed variables.

Climatic variables	Regression		Neural network			
	Model		Root mean square error	Prediction mean error (%)	Root mean square error	Prediction mean error (%)
	Complete	Reduced				
	R <sup>2</sup>	adjusted R <sup>2</sup>				
DCP	0.52**	0.44**	14622.81	11.98	36752.85	152.53
DPPT	0.70**	0.69**	13876.29	126.88	1448.27	12.56
PPT11	0.48**	0.36**	5852.49	-140.15	15417.10	119.06
PPT2	56.40**	0.52**	5201.86	-137.85	10746.46	28.58
PPT811	0.13 <sup>ns</sup>	-	-	-	7525.00	87.78
PPT25	0.50**	0.41**	7409.47	1426.10	9754.85	1458.85
PPT58	0.64**	0.62**	3354.00	796.04	8917.11	375.89

ns, \* and \*\*- not-significant and significant at the levels of 0.05 and 0.01 probability, respectively, by t test (partial regression coefficient) and F test (R<sup>2</sup> and adjusted R<sup>2</sup>), DCP = number of days with precipitation; DPPT = rainfall duration in hours; PPT11 = number of days with rainfall > 11 mm; PPT2 = number of days with rainfall < 2 mm; PPT811 = number of days with rainfall > 8 and < 11 mm; PPT25 = number of days with rainfall > 2 and > 5 mm and PPT 58 = number of days with rainfall > 8 and < 11 mm.

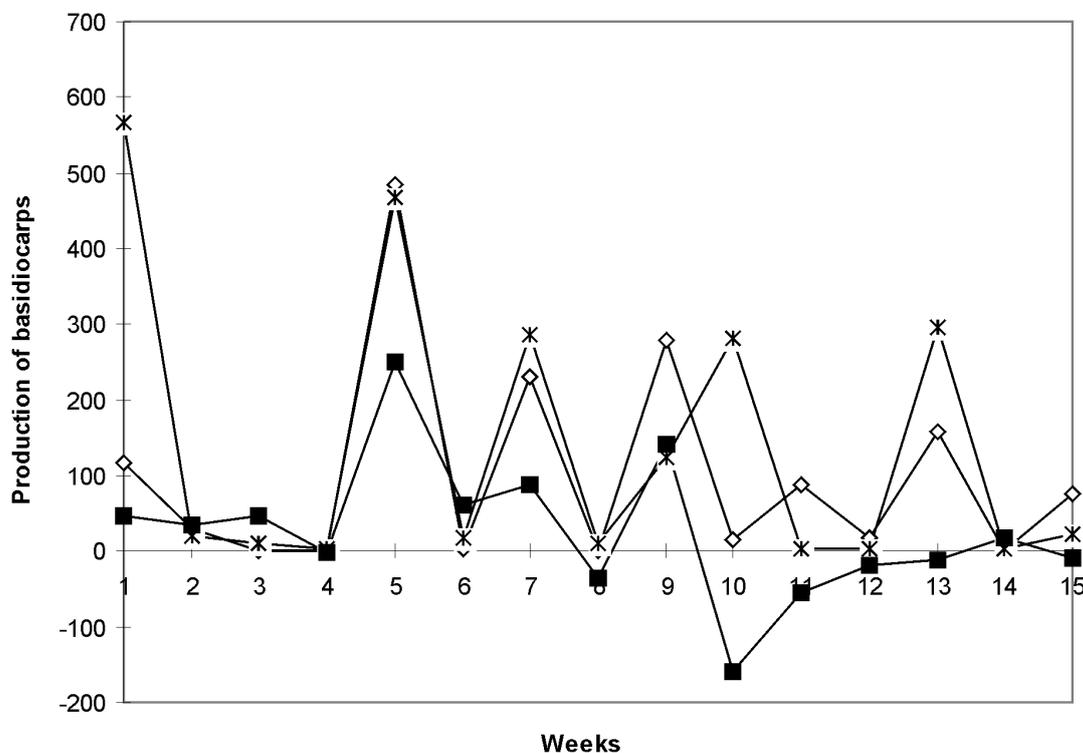
Van der velde (1997) while in the construction; an exhaustive work of variable selection is needed. For instance, Robinson and Mort (1997) have tested several topologies and combined temperature, precipitation, relative humidity,

and wind variables, in the input layer, to build an ANN to forecast frost occurrence in 24 hours. This exhaustive work demands a large number of hours to select the variables, to elaborate topologies, and mainly to train the ANN.

**Table 6:** Basidiocarp production: comparison between regression analysis and neural networks for the week by delayed variable.

S.A.O.F. <sup>1</sup>	Regression				Neural networks	
	Complete R <sup>2</sup>	Reduced R <sup>2</sup> <sub>adjusted</sub>	Root mean square error	Prediction mean error (%)	Root mean square error	Prediction mean error (%)
1	41.95*	29.54**	18557.28	519.93	247167.09	1379.3
2	32.13**	17.34**	6338.43	2279.48	33073.13	1498.15
3	23.05 <sup>ns</sup>	10.0 <sup>ns</sup>	-	-	38863.0	155.79
4	30.51*	21.43 <sup>ns</sup>	83096.48	-823.68	18100.00	1571.69
5	34.56**	31.00**	5135.28	1773.7	2056184.75	29552.72
6	45.15**	42.03**	8998.37	1921.54	42498.02	1921.54
7	41.98**	33.01**	4974.55	1979.37	1762.78	1180.61
8	33.06**	19.12 <sup>ns</sup>	419414.24	5151.32	140387.00	4058.19
9	26.90 <sup>ns</sup>	17.89 <sup>ns</sup>	-	-	98562.56	1056.89
10	22.66 <sup>ns</sup>	14.90 <sup>ns</sup>	-	-	187478.56	2541.56

<sup>1</sup>S.A.O.F. - Weeks earlier observation of fruiting of the pathogen. ns, \* and \*\* - not-significant and significant at the levels of 0.05 and 0.01 probability, respectively, by t test (partial regression coefficient) and F test (R<sup>2</sup> and adjusted R<sup>2</sup>).



**Figure 5:** Production of basidiocarps: the observed values versus predicted values by neural networks ( $\Delta$ -) and regression ( $-*$ -), for the variable DPPT, using the network topology 10-5-2-1.

The time of training varied from 1 second to 1 hour and 9 minutes, totaling approximately 34 hours. Although the time of construction is long, compared to the regression analysis, the NN represent a new approach, in trying to describe witches' broom epidemics and several other phenomena, with which man has not been successful using mathematical modeling. Witches' broom represents one of these phenomena, probably due to genetic diversity of the cocoa trees. This diversity, present in the experiment of Costa (1993), can become one of the variables to be included in the ANN construction. Finally, the use of ANN to describe epidemics is promising and should be an alternative tool for biological data and systems where a great number of variables exist.

In general, the MEF and MSE in the ANN built with two intermediary layers were smaller when compared to that of a layer. However, expand the layer of artificial neurons from one to two, increased the MEF. Neuronal networks with two intermediary layers is less found in the literature. Honjo and Takakura (1991) successfully developed an ANN to forecast supply of water and nutrients for tomato crop using temperature, humidity, and solar radiation as inputs and two intermediary layers. Robinson and Mort (1997) also obtained good results with two intermediary layers for frost forecast. The use of two intermediary layers can constitute an alternative to forecast basidiocarp production and other phenomena that require for the input layer a large number of climatic variables.

The increase in the number of inputs caused an increase in the number of artificial neurons of the intermediary layers. This fact is common when constructing ANN's, as the increase in the number of inputs elevates the complexity in getting an answer (Hertz et al., 2018). The human brain also detects this complexity. To solve problems where two or more variables are involved or in situations that require different senses (vision, touch, and sense of smell), large amount of neurons and different areas in the brain are required, which increases the time of answer. For example, it is easier to solve for the value of  $10+10$  than for  $132.5^2$ . The ANN simulates this behavior and they increases the number of artificial neurons in the intermediary

layers (Movshon, 1996; Van der velde, 1997). The largest time for training nets to describe basidiocarp production reflects this situation.

Some researchers Arizmendi et al. (1993); Silva and Silva (1995); Paul and Munkvold (2016) have successfully used temporal series to forecast some phenomena. These series are also frequently used to forecast multinational stock values and agricultural prices. However, for basidiocarp production, the results with temporal series were not good. For instance, when the output was zero, the tendency was to assume high values to forecast the production. Probably, the number of observations was small to describe this phenomenon. The graphs of the temporal series for both basidiocarp production and disease intensity suggested a nonlinear, non-periodical or chaotic system, because the variables did not exhibit a pattern of behavior in the observed interval (Bianchi; Arizmendi and Sanchez, 1992; Arizmendi et al. 1993). Therefore, to obtain better results, it is required either a larger number of observations or the construction of new ANN in intervals of time capable to describe the trends.

Regression models and ANNs were more appropriate to describe disease intensity than basidiocarp production. For the models chosen in the regression analysis, the variables that best explained disease intensity were T25 and NDR11, although for Costa (1993) Tmax was the best explanatory variable. Costa (1993) observed high  $R^2$  values for the weeks 8 and 9 before disease evaluation. Smallest MSE's and MEF values were found for all the regression models and adjusted ANNs. Disease intensity was the dependent variable with larger possibility of use in the description/forecast of witches' broom. ANN seems to be a good alternative for intensity of cocoa witches' broom, as lower MEF and MSE were generated with it than with regression analysis. With greater experience and training of the net, smaller mistakes are made. Therefore, to refine the technology of ANN, more studies are required with a larger data set obtained in several places with climatic variables.

A criticism to the ANN is the black box nature of the results (Hertz et al. 2018), because it is not well known how the nets arrive to a result. Additionally, comparative variables do

not exist to analyze ANN suitability, as it is the case of the  $R^2$ , the test F and the test t, among other, to test statistical models. However, it is not well known either, how the human brain obtains an answer. In addition, there is no mathematical explanation, as proposed by statistics, but the hypothesis biological-electro-chemistry. If the ANNs are proposed to simulate the human brain, then, it is fair to consider that they can imitate its behavior. Certain people can successfully foresee the weather due to the accumulation of experience and several observations of the nature that surrounds them along the years. Similarly, the neural nets may also simulate this behavior.

## REFERENCES

- ALVES, D. P. et al. Artificial neural network for prediction of the area under the disease progress curve for tomato late blight. **Scientia Agrícola**, 74(1):51-59, 2017.
- ARANZAZU, H. F. Epidemiologia de la escoba de bruja enfermedad del cacao en la zona marginal cafetera y Llanos Orientales de Colombia. **Informe de Resultados Instituto Colombiano Agropecuario** (36). Colombia, 1990.
- ARIZMENDI, C. M. et al. Time series predictions with neural nets: application to airborne pollen forecasting. **International Journal of Biometeorology**, 37(3):139-144, 1993.
- BASTOS, C. N.; SILVA, H. M. Doenças do cacauero na Amazônia Brasileira. **Comunicado Técnico Especial CEPLAC** (2). Belém: CEPLAC/DEPEA/COPEs, 1980. 42 p.
- BASTOS, C.N. Epifitologia, hospedeiro e controle de vassoura-de-bruxa *Crinipellis perniciosa* (Stahel) Singer. **Boletim Técnico CEPLAC** (168). Ilhéus: CEPLAC/CEPEC, 1990. 21p.
- BASTOS, C.N. Estudos desenvolvidos e perspectivas de controle biológico da vassoura-de-bruxa do cacauero. **Comunicado Técnico CEPLAC** (24). Belém: CEPLAC, 1991. 4p.
- BIANCHI, M. M.; ARIZMENDI, C. M.; SANCHEZ, J. R. Detection of chaos: New approach to atmospheric pollen series analysis. **International Journal of Biometeorology**, 36(3):172-175, 1992.
- COLLINS, M. Empiricism strikes back: neural networks in biotechnology. **Nature Biotechnology**, 11:163-166, 1993a.
- COLLINS, M. Practical applications of neural nets fermenter control. **Binary**, 5(1):11-16, 1993b.
- COSTA, J.C.B. **Progresso da vassoura de bruxa em órgãos vegetativos do cacauero em Altamira e Tomé-Açu, Pa**. Viçosa: Imprensa Universitária, 1993. 122p
- COSTA, J.C.B. et al. Indução de resistência em mudas de cacauero contra *Moniliophthora perniciosa* por produto à base de mananoligossacarídeo fosforilado. **Tropical Plant Pathology**, 35(5):285-294, 2010.
- DAS, K.; EVANS, M. D. Detecting fertility of hatching eggs using machine vision II: Neural network classifiers. **Transactions of the ASAE**, 35:2035-2041, 1992.
- EBERHART, R.C.; DOBBINS, R.W. **Neural network PC tools. A practical guide**. 1.ed. San Diego: Academic Press, 1990. 440p.
- FAO, Food and Agriculture Organization of the United Nations. **FAO statistical pocketbook**. Rome: FAO, 2015. 326p.
- GOODRIDGE W. et al. Intelligent diagnosis of diseases in plants using a hybrid Multi-Criteria decision making technique. **Computers and electronics in agriculture**, 133:80-87, 2017.
- HERTZ, J.; KROGH, A.; PALMER, R. G. **Introduction to the theory of neural computation**. Boca Raton: CRC Press, Taylor & Francis Group, 2018. 321p.
- HONJO, T.; TAKAKURA, T. **Identification of water and nutrient supply to hydroponic tomato plants by using neural nets**. In: Hashimoto Y.; Day W. Mathematical and control applications in agriculture and horticulture. Oxford: Pergamon Press, 1991. v.1, p.285-288.
- LIU, Z. Y.; WU, H. F.; HUANG, J. F. Application of neural networks to discriminate fungal infection levels in rice panicles using hyperspectral reflectance and principal components analysis. **Computers and Electronics in Agriculture**, 72:99-106, 2010.



MOVSHON, J. A. Using computation to help relate neurobiology and behavior. *Network: Computation in neural systems*, 7:157-159, 1996.

NATER, E. A.; NATER, K. D.; BAKER J. M. Application of artificial neural system algorithms to image analysis of roots in soil I. Initial results. *Geoderma*, 53(3-4):237-253, 1992.

PINTO, A. C. S. et al. Descrição da epidemia da ferrugem do cafeeiro com redes neuronais. *Fitopatologia Brasileira*, 27(5):517-524, 2002.

ROBINSON, C.; MORT, N. A neural network system for the protection of citrus crops from frost damage. *Computers and Electronics in Agricultural*, 16(3):177-187, 1997.

RUMELHART, D. E.; WIDROW, B.; LEHR, M. A. The basic ideas in neural networks. *Communications of the ACM*, 37(3):87-91, 1994.

SILVA, C. A. B.; SILVA, E. B. Previsão da recepção de leite em uma usina laticinista: uma aplicação de redes neurais. *Revista de Economia e Sociologia Rural*, 33:89-97, 1995.

VAN DER VELDE, F. On the use of computation in modelling behavior. *Network: Computation in Neural Systems*, 8(1): 1-32, 1997.