Estimation of mortality and survival of individual trees after harvesting wood using artificial neural networks in the amazon rain forest

Leonardo Pequeno Reis[a,*], Agostinho Lopes de Souza[b], Pamella Carolline Marques dos Reis[b], Lucas Mazzei[c], Carlos Pedro Boechat Soares[b], Carlos Moreira Miquelino Eleto Torres[b], Liniker Fernandes da Silva[d], Ademir Roberto Ruschel[c], Lyvia Julienne Sousa Rêgo[b], Helio Garcia Leite[b]

[a] *Instituto de Desenvolvimento Sustentável Mamirauá, Grupo de pesquisa em Ecologia Florestal, Estrada do Bexiga, 2.584, Bairro Fonte Boa, Cx. Postal 38 69.553-225, Tefé, AM, Brazil*
[b] *Universidade Federal de Viçosa, Campus Universitário de Viçosa, Departamento de Engenharia Florestal, Avenida Peter Henry Rolfs, s/n, CEP 36.570-900, Viçosa, MG, Brazil*
[c] *Embrapa Amazônia Oriental, Travessa Doutor Enéas Pinheiro, CEP 66.095-903, Belém, PA, Brazil*
[d] *Universidade Federal do Recôncavo da Bahia, R. Rui Barbosa, n. 710, CEP 44380-000, Cruz das Almas, BA, Brazil*

ARTICLE INFO

ABSTRACT

Modeling individual trees in tropical rain forests in the Amazon allows for the safe use of scarce resources in a sustainable way. Unfortunately, in the Brazilian Amazon, rain forest growth and production models are not yet used to estimate future forest stock. Thus, forest management plans do not present technical-scientific support that guarantees sustainable production of wood throughout the cutting cycle. Therefore, this work aims to estimate the survival and mortality of individual trees in a selectively harvested forest using Artificial Neural Networks (ANN) to support silvicultural decisions in forest management in the Amazon rain forest. In 1979, a selective harvest was carried out, with 72.5 m3 ha-1 in an area of 64 ha in Floresta Nacional do Tapajós, in the state of Pará, Brazil. In 1981, 36 permanent plots were installed at random and inventoried. Nine successive measurements were carried from 1982 to 2012. In the modeling, classification, survival, and mortality, training and ANN testing were performed, using input variables such as: different semi-distance-independent competition indices (*DSICI*), diameter measured (*dbh*), forest class (*FC*), trunk identification class (*TIC*), competition index (*CI*), growth groups (*GG*), liana infestation intensity (*liana*); and crown lighting (*CL*); Damage to tree (*D*) and tree rotting (R). The categorical output variables (Classification) were Dead or Surviving tree. Overall efficiency of the classification was above 89% in training and above 90% in the test for all ANNs. Survival classification hit rate was above 99% in the test and training for all ANNs but the mortality score was low, with hit rates below 6%. The overall Kappa coefficient was below 8% for all ANNs (ranked "poor") but all ANNs were above 55% in the survival classification (ranked "good"). ANN estimates the individual survival of trees more accurately but this does not occur with mortality, which is a rarer event than survival.

1. Introduction

Studies on the dynamics of rain forests are important to understand the evolution of the forest ecosystem after anthropic disturbances, for example, during forest management for wood production. These studies provide information to model growth and production, and for prognosis on the forest structure throughout the cutting cycle. One of the main contributions is the use of models for individual trees, which is one of the alternatives to manage rain forests with a view to sustainability (Reis et al., 2016).

The individual tree models estimate the survival and mortality, these components of forest dynamics are required for correct prognosis on number of trees, basal area, distribution of diameters and production.

One of the problems in modeling mortality is that several random factors may cause the death of trees. For example, regular mortality is

caused by aging, suppression and competition, as well as events that occur less frequently, for example, normal incidence of plagues and diseases, and meteorological phenomena (droughts, storms, etc.); by comparison, irregular mortality may be caused by large-proportion fires, plague and disease outbreaks, as well as more severe adverse meteorological conditions (Vanclay, 1994). Mortality also occurs as a result of injuries induced by wood harvesting, which may damage roots and barks, creating points of entry of plagues and diseases; as well as disturbances on the canopy, which may lead to disadvantages to some tree species (Vanclay, 1994).

Tree mortality ratio may be reached using regression (Phillips et al., 2004, 2003; Valle et al., 2007). These authors used a system of equations to estimate the likelihood of natural mortality of trees by considering only a diameter-dependent stochastic process, or artificial intelligence methods, mainly Artificial Neural Networks (ANNs) (Diamantopoulou, 2005; Reis et al., 2016). ANNs are computer models inspired by the nervous system of living beings. One ANN creates a set of parallel processing units, characterized by artificial neurons that are interconnected through a large number of interconnections (Silva et al., 2010).

Different studies that modelled mortality and survival of individual trees using ANN found a more adequate fit than traditional statistical techniques (Guan and Gertner, 1991a, 1991b; King et al., 2000). They have shown that it is possible to have a prognosis on the individual survival and mortality of trees using ANN. These authors used two models to estimate tree mortality: one model with two independent variables, DBH and increase in DBH, and one model with three variables, with an extra categorical variable which represented the condition of the tree. The output was categorical (classification), that is, the dead tree was coded with 0 and the survivor with 1. The ANN results were compared with logistic regression, and better responses were found to predict mortality.

This estimate with ANN for the purpose of classification, indicating whether the tree is dead or alive, shows that ANN has a far more complex function than traditional classification techniques. The final discriminant function is highly flexible and non-linear, and it offers better separation (King et al., 2000).

However, only one study was found on modelling of tree mortality in rain forests (Castro et al., 2015), but it did not involve an extended period of time nor was it conducted in a harvested forest, whose dynamics is quite different from unharvested areas (Reis et al., 2015).

Growth prognosis on the individual diameter of trees, after harvesting in a tropical rain forest in the Amazon, was precisely estimated using ANN (Reis et al., 2016), and the same occurred in other types of uneven-aged forests (Ashraf et al., 2015; Richards et al., 2008).

Given the problem in offering a prognosis on the survival and mortality of individual trees in tropical rain forests, the aim of this study was to offer a prognosis for the individual survival and mortality of trees using post-harvest artificial neural networks in the Amazon, in order to offer input for forestry decisions in forest management.

2. Material and methods

2.1. Study area

The study area is located in the Tapajós National Forest, near Km 67 (55° 00′ W, 2° 45′ S) of the BR-163 Highway, Cuiabá-Santarém. It is part of the Amazon biome and the typology is solid-ground, Dense Ombrophilous Forest. The climate of the region is humid and tropical with mean annual temperature of 26 °C, and it is classified as Ami according to Köppen's system. Mean relative humidity corresponds to 86%, with mean annual rainfall from 1900 to 2200 mm. It has flat to wavy topography, with the occurrence of a Dystrophic Yellow Latosol (Alvares et al., 2013; Costa Filho et al., 1980).

In the Tapajós National Forest, especially in the study area, Costa Filho et al. (1980) reported the use of selective harvest, conducted

during the 1940s, for four species with high commercial value: Brazilian rosewood (*Aniba rosaeodora* Duke), Brazilian redwood (*Manilkara huberi* (Ducke) A. Chev.), Brazilian walnut (*Cordia goeldiana* Huber) and cedar (*Cedrela odorata* L.). In 1979, an intensive harvest of 64 wood species was conducted on 64 ha of the study area, with mean extraction volume of 72.5m^3 ha^{-1} (Reis et al., 2010).

The species that stood out in terms of harvest volume, at the time, were: *Hymenaea courbaril* L., *Carapa guianensis* Aubl., *Manilkara huberi*, *Lecythis lurida* (Miers) S. A. Mori., *Bertholletia excelsa* Humb. & Bonpl., *Astronium lecointei* Ducke, *Goupia glabra* Aubl., *Virola michelii* Heckel, *Erisma uncinatum* Warm. and *Terminalia amazonia* (J. F. Gmel) Exell, which, together, represented 47.4% of the total extracted volume (Reis et al., 2010). The harvest was conducted according to two treatments: cutting all trees with *dbh* ≥ 45 cm, on 39 ha; and cutting the trees with *dbh* ≥ 55 cm, on 25 ha (Costa Filho et al., 1980). However, the treatments were considered together, by creating only one community, while taking into account the high similarity found in the comparisons which had been made (Reis et al., 2010).

In 1981, 36 permanent plots of 50 m x 50 m each were randomly installed, where all trees with *dbh* ≥ 5 cm were botanically identified *in loco*. New measurements for these permanent plots occurred in 1982, 1983, 1985, 1987, 1992, 1997, 2007, 2010, and 2012 (Reis et al., 2016).

2.2. Variables and data used for training and testing of neural networks

The permanent plots were divided into two groups: one group consisted of 29 plots for training of ANNs, and one group had 7 plots, for the generalization of trained ANNs, with a total of 80% of data for training and 20% for generalization (test). The plots used in the generalization (test) were not part of the training set. This was to evaluate the model with independent data to the training of ANNs (Reis et al., 2016). A total of 78,067 individuals were monitored over time; there were 8332 cases of mortality and 69,735 cases of survival. For mortality, the training used 6819 trees while the test used 1513. For survival, the training used 56,421 trees while the test used 13,314.

To model the mortality and survival of individual trees, the entry variables were: diameter measured at a height of 1.30 m (*dbh*), forest class (*FC*), trunk identification class (*TIC*), competition index (*CI*), growth groups (*GG*), liana infestation intensity (*liana*): variable not observed; *liana*1: no presence of liana on the tree; *liana*2: presence of lianas, however, with no injuries; and *liana*3: presence of lianas, restricting growth); and crown lighting (*CLI0*: variable not observed; CLI1: emerging top or completely exposed to light; CLI2: partially lighted top, that is, partially covered by neighboring tree tops; CLI3: top completely covered by neighboring tree tops) (Reis et al., 2016); injuries to the tree (D0: variable not observed; D1: tree with no injuries; D2: mild injuries caused by natural causes; D3: mild injuries caused by harvesting; D4: injuries caused by cutting the lianas; D5: severe injuries due to natural causes; D6: severe injuries caused by harvesting; D10: recovered injuries) and tree rotting (R0: variable not observed; R1: no rotting and R2: presence of rotting). The categorical output variables (Classification) were Dead or Surviving tree.

The forest classes (*FC*) were defined according to the methodology suggested by Silva et al., 2005:

1 Mature forest: the sub-plot shows at least one tree whose diameter is equal to or larger than 40 cm
2 Forest under construction: the sub-plot has at least one tree whose diameter is equal to or larger than 10 cm and smaller than 40 cm
3 Clearing: there is an opening on the canopy of at least 50% of the area of the sub-plot and few or no trees with a diameter larger than 10 cm on the sub-plot. When existing, the crowns project themselves outside the limits of the sub-plot.

Trunk identification classes (*TIC*) were defined using the

methodology suggested by Silva et al., 2005:

1 Living standing tree, complete.
2 Living standing tree, no crown, trunk > 4.0 m.
3 Living standing tree, no crown, trunk < 4.0 m.
4 Living fallen tree.
5 Supported tree due to natural cause.
6 Bent tree due to natural cause.
7 Arched tree due to natural cause.

The competition indexes tested on this study were the distance semi-independent competition indexes (*DSICI*), among which are:

$$DSICI_1 = \frac{D_t^2}{\overline{d}^2} \tag{1}$$

Eq. (1) Adapted from (Glover and Hool, 1979)

$$DSICI_2 = Bal_i \tag{2}$$

Eq. (2) Adapted from (Stage, 1973)

$$DSICI_3 = Z_1 \sum_{i=1}^{n1} \frac{D_i}{D_t} + Z_2 \sum_{j=1}^{n2} \frac{D_j}{D_t} \tag{3}$$

Eq. (3) Adapted from Phillips et al. (2004)

where, D_t is the diameter of the study tree; $\overline{d}^2$ is the arithmetic mean of the diameters on the sub-plot of the study tree; Bal_i is the sum of the sectional areas of the neighboring trees larger than the sectional area of the study tree, on the sub-plot; Z_1 and Z_2 are the relative importance coefficients for competition of zones 1 and 2, respectively; D_i and D_j are the "over-topping" trees on both zones. n_1 and n_2 are the total number of "over-topping" trees on the three zones. The "over-topping" trees are the ones whose diameter is larger than that of the study tree on the sub-plots in both zones.

Zona 1 is a 10 × 10 m square containing t trees (Fig. 1). Zone 2 is defined as relative to zone 1. Random weights are attributed, reflecting the competition on each zone, considering $Z_1 = 9$ and $Z_2 = 4$ (Phillips et al., 2004; Reis et al., 2016).

A cluster analysis was made according to the Euclidean distance method in order to classify the species into growth groups by using the annual periodical increment to the diameter (API_{dbh}), in mm year^{-1} (Reis et al., 2016), of the botanical families over the different periods (1981–1982, 1982–1983, 1983–1985, 1985–1987, 1987–1992, 1992–1997, 1997–2007 and 2007–2012), where four growth groups were created (Table 1).

2.3. Training and evaluation of ANNs

The training of ANNs consisted on applying a set of organized steps with the purpose of adjusting the weights and thresholds of the neurons. Therefore, the purpose of such adjustment process, also known as
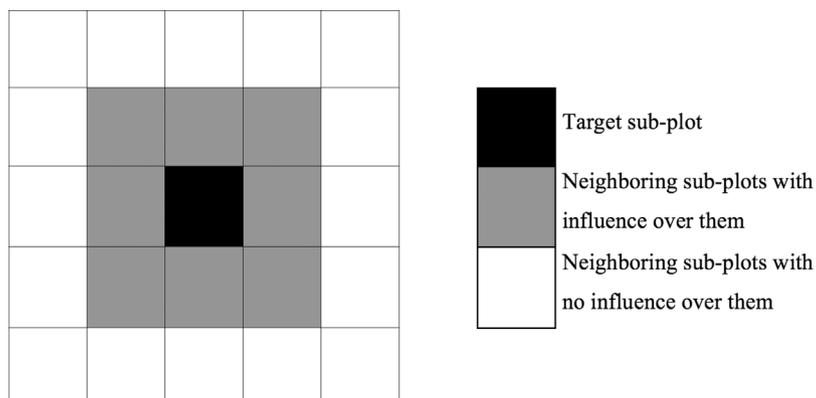
learning algorithm, was to tune the network in order for the responses to be similar to the output values (Silva et al., 2010).

In the modeling process, 1200 ANNs were trained, 300 for each competition index being evaluated (*DSICI₁*, *DSICI ₂*, and *DSICI ₃*) and 300 ANNs without the competition index variable in order to simplify the ANNs in the modeling process (Table 2).

Training was based on the generalized delta rule, also known as the backpropagation algorithm, applied in multi-layer feed forward networks–Multilayer Perceptron (MLP), as shown in Fig. 2.

The backpropagation algorithm is performed by successive applications of two specific phases. According to Silva et al. (2010) the first phase to be applied is called "forward propagation", in which signals {x_1, x_2,…, x_n} of a sample of the training set are inserted into the network inputs and are propagated layer by layer until the respective outputs are produced. The application of this phase aims only to obtain the network responses, taking into account only current values of synaptic weights and thresholds of its neurons, which will remain unchanged during each execution. Subsequently, the responses produced by the outputs of the network are compared with the respective desired responses that are available, the respective deviations (errors) between the desired responses and those produced by the output neurons are then calculated, which will subsequently be used to adjust the weights and thresholds of all their neurons. Because of these error values, the second phase of the backpropagation method, called "reverse propagation" (backward), is then applied. The changes (adjustments) of the synaptic weights and thresholds of all the neurons of the network are executed during this phase (Silva et al., 2010).

In summary, the successive applications of the forward and backward phases make the synaptic weights and thresholds of the neurons automatically adjust at each iteration, implying a gradual decrease in the sum of the errors produced by the responses of the network compared with the desired ones (Silva et al., 2010).

For training, the Intelligent Problem Solver (IPS) tool of the Statistica 13 software (StatSoft Inc, 2016) was used to analyze the activation functions (Table 3) of the intermediate and output layers (Identity, Logistics, Hyperbolic Tangent and Exponential). The initial weights of the networks were randomly generated, and the stopping criterion of the algorithm occurred when the mean square error or the cross-entropy error began to increase; on that occasion the training was interrupted. Network training continued up to 10,000 cycles as long as the error was decreasing.

For the purpose of training, only one hidden layer was used, and the neuron number interval on this layer was defined using the Fletcher-Gloss method (Silva et al., 2010). The neuron number interval was established according to the number of input and output variables, considering the following expression:

$$(2.\sqrt{n} + n_2) \leq n_1 \leq (2.n + 1) \tag{4}$$

where n is the number of network inputs, n_1 is the amount of neurons on the hidden layer, and n_2 is the amount of neurons on the output

Fig. 1. Competition zones considering a tree on the sub-plot. Adapted from Phillips et al. (2004).

Target sub-plot

Neighboring sub-plots with influence over them

Neighboring sub-plots with no influence over them

Table 1
Growth groups created from the Annual Periodical Increment on *dbh* (API_{dbh}) of the botanical families after the forestry harvesting.

Groups	Families	API_{dbh} (mm year^{-1})
1	To be identified, Chrysobalanaceae, Connaraceae, Lamiaceae, Lecythidaceae, Myrtaceae, Ochnaceae, Opiliaceae, Quiinaceae, Rubiaceae, Rutaceae, Sapindaceae, Siparunaceae and Violaceae.	1.355
2	Achariaceae, Annonaceae, Apocynaceae, Aquifoliaceae, Boraginaceae, Burseraceae, Caryocaraceae, Celastraceae, Clusiaceae, Dichapetalaceae, Ebenaceae, Elaeocarpaceae, Euphorbiaceae, Hippocrateaceae, Lacistemaceae, Lauraceae, Leguminosae-papilionoideae, Malvaceae, Melastomataceae, Moraceae, Myristicaceae, Nyctaginaceae, Olacaceae, Polygonaceae, Salicaceae, Sapotaceae, Solanaceae and Ulmaceae.	2.731
3	Anacardiaceae, Combretaceae, Goupiaceae, Humiriaceae, Leguminosae-caesalpinioideae, Malpighiaceae, Meliaceae and Simaroubaceae.	3.856
4	Araliaceae, Bignoniaceae, Bixaceae, Caricaceae, Leguminosae-mimosoideae, Rosaceae, Urticaceae and Vochysiaceae.	7.771

Table 2
Variables used to train the artificial neural networks (ANN) in a forest after forestry harvesting in the Eastern region of the Amazon.

| Competition index	Input variables	Number of trainings	Categorical output
$DSICI_1$	dbh_1, FC, GG, TIC, liana, CL, $DSICI_1$, D, R	300	D and S
$DSICI_2$	dbh_1, FC, GG, TIC, liana, CL, $DSICI_2$, D, R	300	D and S
$DSICI_3$	dbh_1, FC, GG, TIC, liana, CL, $DSICI_3$, D, R	300	D and S
No DSICI	dbh_1, FC, GG, TIC, Liana, CL, D, R	300	D and S
General Total		1200	

Continuous input variables: *DSICI*: Distance semi-independent competition index; dbh_1: Current diameter measured at 1.30 m from the soil (mm); Categorical input variables: *FC*: forest class; *GG*: growth group; *TIC*: trunk identification class; *liana*: liana infestation intensity; *CL*: crown lighting; D: damage and R: rotting. Categorical output: Dead Tree (D) or Surviving Tree (S).

layer.

The maximum defined by this method avoids the memorization of input data (overfitting) or the insufficient collection of information during training (underfitting).

In order to observe whether there is any simple linear correlation among the continuous input variables and the ANN output, a simple linear correlation statistical analysis was made at 5% of significance, using the mean of the permanent plots. With samples of *n* pairs of values X_i, Y_i (i = 1, 2, ..., *n*), the (r_{xy}), correlation may be calculated from the following expression:

$$r_{YX} = \frac{Cov(Y, X)}{\sqrt{S^2(X)S^2(Y)}} \tag{5}$$
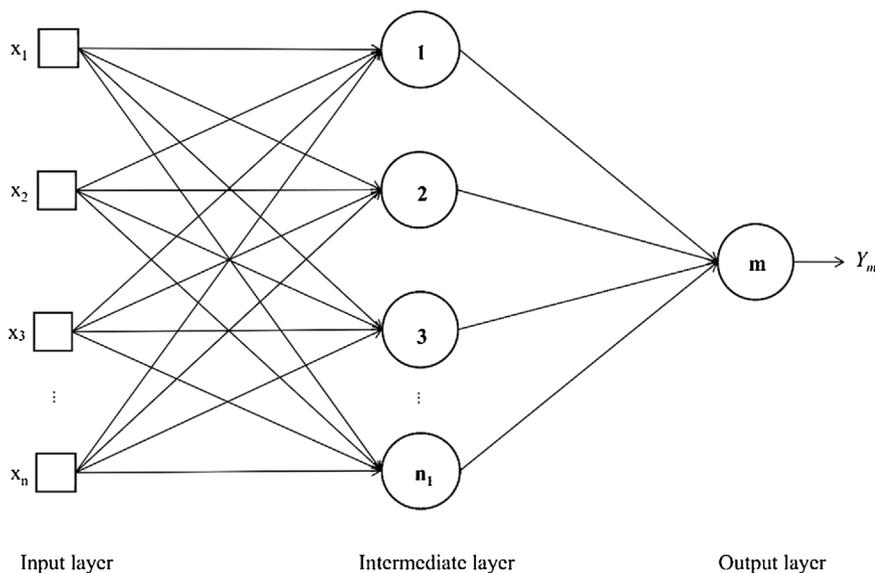
Table 3
Activation functions used in the training of artificial neural networks.

Function	Equation	Description	Interval
Identity	α	With this function, the activation level is transmitted directly as the output of the neurons.	$(-\infty, +\infty)$
Logistic	$\frac{1}{1+e^{-\alpha}}$	This is an S-shaped curve (sigmoid).	$(0, 1)$
Hyperbolic tangent	$\frac{e^{\alpha} - e^{-\alpha}}{e^{\alpha} + e^{-\alpha}}$	It is a symmetric S-shaped (sigmoid) function.	$(-1, +1)$
Exponential	$e^{-\alpha}$	Negative exponential function.	$(0, +\infty)$

α: is the net input of a neuron, for MLP $-$ is the weighted sum of the inputs of the neurons.

The classification was evaluated by generating the error matrix (2 × 2), and omission error (OE), commission error (CE), hit rate, global accuracy Index and total Kappa coefficient (Cohen, 1960) and conditional Kappa coefficient (by class) could be calculated. The Kappa statistics is usually used to evaluate the accuracy of classification (King et al., 2000; Moisen and Frescino, 2002):

$$k = \frac{N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} (x_{i+}. x_{+i})}{N^2 - \sum_{i=1}^{r} (x_{i+}. x_{+i})} \tag{6}$$

$$k_{cond} = \frac{Nx_{ii-}(x_{i+}. x_{+i})}{Nx_i - (x_{i+}. x_{+i})} \tag{7}$$

where:

k = Kappa coefficient; k_{cond} = conditional Kappa coefficient by class; *r* = number of rows on the matrix; x_{ii} = number of observations on the row(*i*) and column(*i*), diagonal values of the matrix; x_{i+} = total marginal of the row (i); x_{+i} = total marginal of the column (i); and

Fig. 2. A multi-layer feedforward network consisting of an input layer composed of n signals, a hidden neural layer, consisting of n_1 neurons and an output neural layer composed of *m* neurons representing the respective output values of the application. Adapted from Silva et al. (2010).

Table 4

Linear correlation of Pearson between input and output variables used in ANN training and test. S (%): Average survival rate; M (%): Average mortality rate; *DSICI*: semi-distance-independent competition index; dap_1: diameter at the beginning of t time period of the object tree (mm).

Variables	S (%)	M (%)
$DSICI_1$	− 0.0379	0.0379
$DSICI_2$	0.1635	− 0.1635
$DSICI_3$	0.0926	− 0.0926
dbh	0.2930	− 0.2930

*p < 0.05.

N = total number of observations.

The Kappa coefficient may be evaluated using the following classification quality scale (Landis and Koch, 1977): $k \leq 0.20$ (poor), 0.21–0.40 (average), 0.41–0.60 (moderate), 0.61–0.80 (satisfactory) and 0.81–1.00 (very satisfactory).

The diametric distribution of the surviving trees was evaluated for the different years using the Chi-square adherence test (χ^2) at 5% significance.

3. Results

The means of the continuous input independent variables used on ANN were not linearly correlated with the mean relative to the survival and mortality rates (Table 4). No competition index tested showed any relationship to mortality and survival percentages.

Global efficiency of the classification was over 89% during training and over 90% for the test, for all ANNs (Table 5). The hit rate for classification of the number of surviving trees was over 99% for the test and training for all ANNs. This percentage was considered to be satisfactory for classification of survival. However, the classification for number of dead trees showed low accuracy, with hit rates below 6%, for the test and the training of ANNs. Nevertheless, the accuracy of the test was higher than that of training, and this characterizes an efficient generalization of ANN for independent data.

The ANN that used $DSICI_3$ was the one with the highest hit rate for mortality, and the ANN which did not use any competition index had the lowest hit rate − under 5% for mortality. This ANN showed values that were very similar, regarding hit rate for the survival classification, with 99.72% for the classification, and 90.05% for global efficiency, which was similar to the efficiency rate of ANNs which used $DSICI_1$ and $DSICI_3$.

All ANNs showed an omission error (OE) over 94% for the training and the test regarding the mortality classification; in opposition, the OE for the survival classification was below 0.4%. This error represents that trees are omitted from their correct class and they are assigned to another class (Table 6). The commission error (CE) for all ANNs on the training and the test were below 38% and 65%, respectively, for survival and mortality. This error represents the trees that were mistakenly included in a certain class when they should be assigned to another

class.

The general Kappa coefficient was below 8% for all ANNs (Table 6), both for the training and the test. According to the classification quality scale of the Kappa coefficient (Landis and Koch, 1977), the classification of all ANNs was considered as "poor" ($k \leq 0.20$). The network that showed the highest Kappa coefficient was the one that used $DSICI_3$, and the lowest one was the one that did not use any competition index.

As regards the Kappa coefficient calculated by class (conditioned), all ANNs reached above 55% for survival (Table 6). A highlight was the ANNs that used the competition indexes $DSICI_2$ and $DSICI_3$, which achieved Kappa coefficients of 64.2 and 61.1%, respectively, for the test data. These classifications were considered as "satisfactory", according to the classification quality scale of the Kappa coefficient (0.61–0.80).

The mortality classification on all networks showed a Kappa coefficient below 5% (Table 6), both for the test and training. The ANN that showed the highest Kappa coefficient was the one that used $DSICI_3$, for training (3.43%) and the test (4.50%), according to the classification quality scale of the Kappa coefficient, the classification of this network was also considered as "poor" ($k \leq 0.20$).

The exponential diameter distribution of the estimated and observed data of the surviving trees showed an inverted-J shape (Fig. 2). The estimated distribution was not different from the distribution observed (Table 6) for the years 1982, 1983, 1985, 1987 and 2012 for the test data across all ANNs. The estimated diametric distribution for the years 1992, 1997 and 2007 was different from the observed distribution (Table 7); during those years, survival was overestimated at the smallest diameter class center (7.5 cm), for all networks, for the test data (Fig. 3).

4. Discussion

The high global efficiency indicated that the quality of the classification was efficient; however, this efficiency considers only the main diagonal of the error matrix, and it does not include the omission and contingency errors for each one of the classes (Tables 5 and 6). This may be observed by the high hit rate for survival, over 99%, and the fact that mortality did not exceed 6%.

Since the Kappa coefficient uses the entire error matrix, not only the main diagonal of this matrix, it offers a better representation of the classification quality (Table 6). The general classification of the mortality and survival was not precise, and it was classified as "poor"; this is mainly due to the mortality classification, with the Kappa coefficient also classified as "poor".

The lack of a linear correlation across the means of the continuous input variables and the survival and mortality percentages may indicate that such means did not represent the complex and stochastic relationship of the dynamics of these rates; however, a non-linear correlation may exist across the variables (Table 4). The means of the tree diameters (Vanclay, 1994) and the competition indexes by plot were not linearly correlated with mortality; however, they are widely used as indicators for the modelling of individual trees (Phillips et al., 2004).

Table 5

Training precision measures and Artificial Neural Network (ANN) test in the classification of survival (S) and mortality (M) of individual trees. *ICSID*: semi-distance-independent competition index; MLP: multilayer perceptron.

ANN	Index	Architecture MLP					Training			Test		
		Numb. input	Numb. Neurons Intermediate layer	Numb. Neurons-Output layer	Activation functions Intermediate layer	Activation functions Output layer	HR S	HR M	OEE	HR S	HR M	OEE
$DSICI_1$	2	35	44	01:01	Logistic	Exponential	99.71	3.92	89.38	99.7	5.22	90.05
$DSICI_2$	1	35	16	01:01	Hyperbolic tangent	Hyperbolic tangent	99.71	4.03	89.39	99.7	5.02	90.07
$DSICI_3$	3	35	66	01:01	Hyperbolic tangent	Exponenti al	99.73	4.09	89.42	99.7	5.29	90.05
No DSICI	5	34	59	01:01	Logistic	Hyperbolic tangent	99.73	3.77	89.38	99.7	4.89	90.05

HR: Hit rate (%); OEE: Overall equipment effectiveness (%).

Table 6

Errors and precision of training and Artificial Neural Networks (ANN) in the classification of survival and mortality of individual trees. EO: Error of omission (%); EC: Error of commission (%); General and conditioned Kappa coefficient (%); *ICSID*: semi-distance-independent competition index.

ANN	Training							Test						
	EO (%)		EC (%)		Cond. K. (%)		General K.	EO (%)		EC (%)		Cond. Kappa (%)		General. K (%)
	S	M	S	M	S	M		S	M	S	M	S	M	
DSICI 1	0.29	96.08	10.43	37.76	57.67	3.26	6.17	0.31	94.78	9.75	34.17	62.0	4.45	8.30
DSICI 2	0.29	95.97	10.42	37.36	58.13	3.36	6.36	0.27	94.98	9.77	32.14	64.2	4.30	8.06
DSICI 3	0.27	95.91	10.41	35.42	60.30	3.43	6.49	0.32	94.71	9.75	34.96	61.1	4.50	8.37
No *DSICI*	0.27	96.23	10.44	37.16	58.34	3.14	5.96	0.28	95.11	9.78	33.33	62.9	4.17	7.83

Table 7

Chi-square test () of adherence between the diameter distribution observed and estimated in the years after the forest harvest. *ICSID*: semi-distance-independent competition index.

Years	*DSICI 1*		*DSICI 2*		*DSICI 3*		No *DSICI*	
	Chi-Square (χ^2)	p-value	Chi-Square (χ^2)	p-value	Chi-Square (χ^2)	p-value	Chi-Square (χ^2)	p-value
1982	0.270	0.9999	0.254	0.9999	0.327	0.9999	0.324	0.9999
1983	1.076	0.9998	1.076	0.9998	1.025	0.9998	1.092	0.9998
1985	4.361	0.9296	4.361	0.9296	4.138	0.9409	4.397	0.9277
1987	9.006	0.5315	9.074	0.5251	8713	0.5595	9.155	0.5595
1992	75.234	< 0.0001	75.961	< 0.0001	75.147	< 0.0001	75.961	< 0.0001
1997	27.799	0.0019	28.318	0.0016	27.287	0.0023	29.265	0.0011
2007	102.869	< 0.0001	104.172	< 0.0001	102.869	< 0.0001	104.576	< 0.0001
2012	9.595	0.4767	9.695	0.4676	9.835	0.4551	10.219	0.4215

Degrees of freedom: 10.

The network that used *DSICI₃* showed the highest hit rate for mortality and survival, and the highest Kappa coefficients (Tables 5 and 6). This index is a better representation of competition. It considers competition both within the plot of the study tree as well as in the neighboring plots (Fig. 1), because it uses the competing trees that are larger than the study tree and it shows weights that takes into account the reduction of competition as the study tree gets farther away.

The study forest showed the exponential diameter distribution throughout 31 years, as an inverted-J shape, typical of tropical forests, a trend that was followed by all ANN (Fig. 3). This trend is common in forests with no intensive disturbances, with a larger number of individuals in the smallest diameter classes, which supports this negative exponential shape (Rubin et al., 2006). That indicates that the density of the smaller trees supports the population of larger trees (Rubin et al., 2006).

Non-adherence between the estimated and the observed diametric distribution during the years 1992, 1997 and 2007 (Table 7) occurred as a result of overestimation of the first diameter class, on all tested networks, reaching a difference of 7.5 cm and 29.7, 15.8 and 32.8%, in 1992, 1997 and 2007 respectively, in comparison with *DSICI₃* modelling. That occurred because the trees with smaller diameters are the ones that suffer the effects of competition the most (Farrior et al., 2016; Reis et al., 2016). A natural reduction in survival occurred during those years for this class (7.5 cm) and there was a recovery at the end of the period (2012), which was followed by ANNs (Fig. 2).

The survival rate was followed by all ANNs throughout time after forest harvesting (Fig. 4). With the overestimation on the number of surviving trees for 1992, 1997 and 2007, and with the survival increase in the last period, all ANNs showed the same trend after 33 years of forest harvesting. However, by analyzing the annual survival rate, all tested models followed the trend of the observed data, first with the increase of survival after harvesting, and then with a continuous reduction of survival; that was only possible because of the high hit rate of neural networks in the classification of survival.

The rise on the survival rate at the beginning of the period after the forest harvesting was due to the opening in the canopy, which increased tree growth, while the reduction in the survival rate throughout time

was caused by the continuous closing of the canopy (Reis et al., 2016, 2015).

Although the continuous input variables of the networks did not show any correlation with survival and mortality percentages, their hit rate was due to the categorical variables, as shown by Reis et al. (2016), which had a significant influence on growth, for example, crown lighting (Gustafsson et al., 2016), liana infestation (Campanello et al., 2007; Reis et al., 2015); forest class, represented by the different canopy sizes (Stan and Daniels, 2014); growth groups (Azevedo et al., 2008; Valle et al., 2007) and damage and rotting, which may be caused or may not be derived from harvesting activities that also contribute to mortality in forest stands (Vanclay, 1994), thus corroborating the use of these variables for the modelling of uneven-aged forests.

The problem in modeling mortality in tropical rain forests is related to the stochasticity of this component, which makes the prognosis difficult to establish. That occurs because of the long lifecycle of the trees, which makes mortality a rare event (King et al., 2000). Its description requires a regular survey on a large number of trees (Wernsdörfer et al., 2008) and a series of ecological data that influence regular mortality and which are hard to obtain and are not measured in permanent plots, for example, variables relative to the microclimate of the plots. On the other hand, since it is not such a rare event, and with a higher number of occurrences, the classification of survival was considered as "satisfactory" according to the Kappa coefficient on all tested networks (Table 6).

5. Conclusion

Artificial neural networks are efficient in offering a prognosis on the survival of individual trees after forest harvesting, with hit rates over 99% and Kappa coefficient over 55%. Also, the use of ANNs is indicated to offer a prognosis on the annual survival rate and the diameter distribution at the end of the cutting cycle, therefore, in harvested forests of the Amazon.
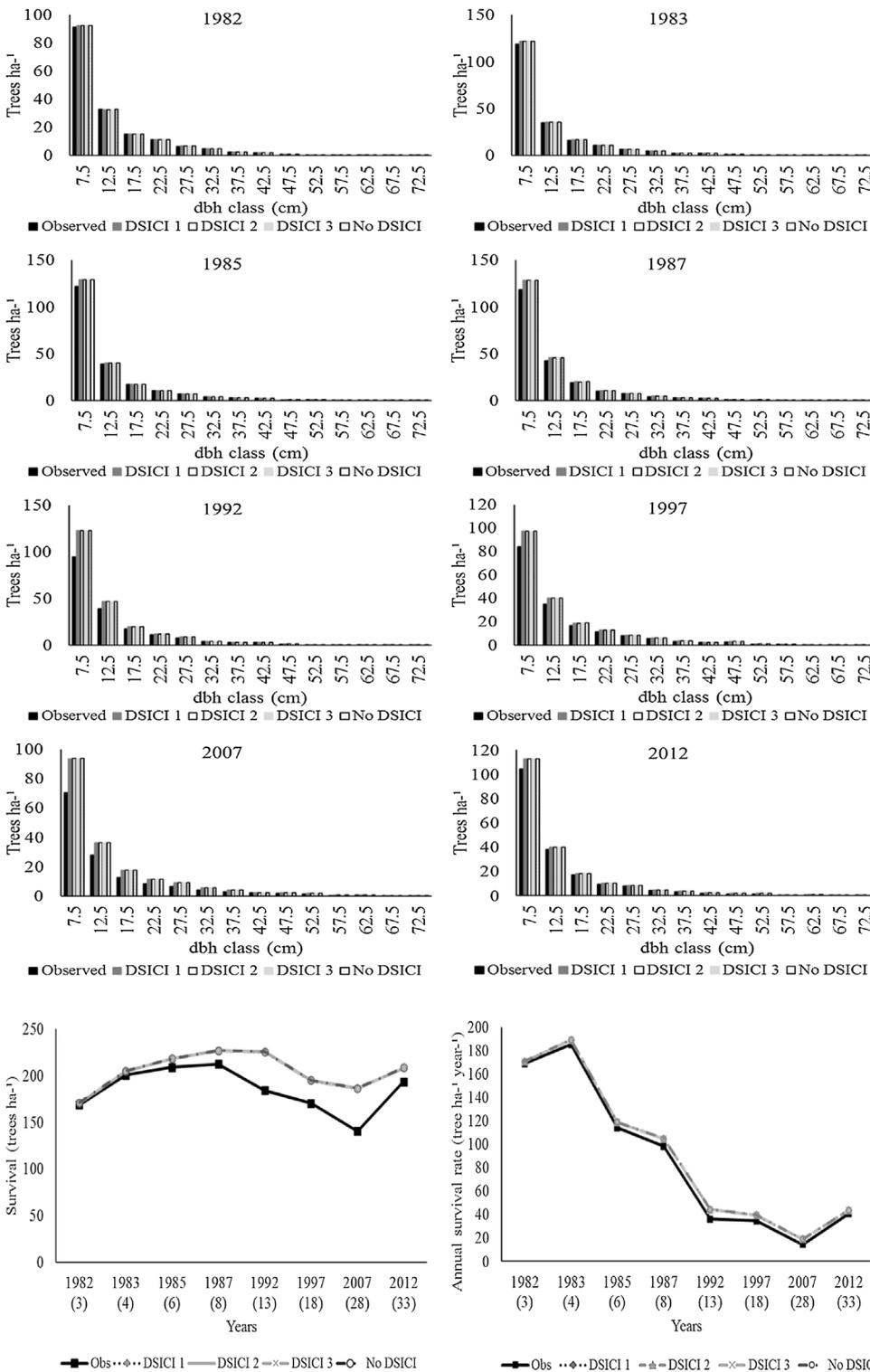
Fig. 3. Observed and estimated diametric distribution by ANN of surviving trees in the years after forest harvest. *ICSID*: semi-distance-independent competition index.

Fig. 4. Survival (trees ha^{-1}) and annual survival rate (tree ha^{-1} year^{-1}) observed and estimated over time after harvesting wood. Values in parentheses represent the years after forest harvest.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ecoleng.2017.12.014.

References

Alvares, C.A., Stape, J.L., Sentelhas, P.C., De Moraes Gonçalves, J.L., Sparovek, G., 2013. Koppen's climate classification map for Brazil. Meteorol. Z. 22, 711–728. http://dx. doi.org/10.1127/0941-2948/2013/0507.

Ashraf, M.I., Meng, F.-R., Bourque, C.P., MacLean, D.A., 2015. A novel modelling approach for predicting forest growth and yield under climate change. PLoS One 10, e0132066. http://dx.doi.org/10.1371/journal.pone.0132066.

Azevedo, C.P., De, Sanquetta, Silva, C.R., Machado, J.N.M., Souza, S.A., Oliveira, C.R., De, M.M., 2008. Simulação de estratégias de manejo florestal na Amazônia com o uso

do modelo SYMFOR. Acta Amazon 38, 51–70. http://dx.doi.org/10.1590/S0044-59672008000100007.

Campanello, P.I., Garibaldi, J.F., Gatti, M.G., Goldstein, G., 2007. Lianas in a subtropical Atlantic forest: host preference and tree growth. For. Ecol. Manage. 242, 250–259. http://dx.doi.org/10.1016/j.foreco.2007.01.040.

Castro, R.V.O., Soares, C.P.B., Leite, H.G., Souza, A.L., Martins, F.B., Nogueira, G.S., Romarco, M.L.O., 2015. Artificial neural networks effectiveness to estimate mortality in a semi-Deciduous seasonal forest. Aust. J. Basic Appl. Sci. 9, 435–444.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37–46. http://dx.doi.org/10.1177/001316446002000104.

Costa Filho, P., Costa, H., Aguiar, O., 1980. Exploraçao mecanizada da floresta tropical úmida sem babaçu Embrapa-CPATU. Circ. técnico 9, 29.

Diamantopoulou, M.J., 2005. Artificial neural networks as an alternative tool in pine bark volume estimation. Comput. Electron. Agric. 48, 235–244. http://dx.doi.org/10.1016/j.compag.2005.04.002.

Farrior, C.E., Bohlman, S.A., Hubbell, S., Pacala, S.W., 2016. Dominance of the suppressed: power-law size structure in tropical forests. Science 351, 155–157. http://dx.doi.org/10.1126/science.aad0592.

Glover, G.R., Hool, J.N., 1979. A basal area ratio predictor of loblolly pine plantation mortality. For. Sci. 25, 275–282.

Guan, B.T., Gertner, G., 1991a. Modeling red pine tree survival with an artificial neural network. For. Sci. 37, 1429–1440.

Guan, B.T., Gertner, G., 1991b. Using a parallel distributed processing system to model individual tree mortality. For. Sci. 32, 871–885.

Gustafsson, M., Gustafsson, L., Alloysius, D., Falck, J., Yap, S., Karlsson, A., Ilstedt, U., 2016. Life history traits predict the response to increased light among 33 tropical rainforest tree species. For. Ecol. Manage. 362, 20–28. http://dx.doi.org/10.1016/j.foreco.2015.11.017.

King, S.L., Bennett, K.P., List, S., 2000. Modeling noncatastrophic individual tree mortality using logistic regression, neural networks, and support vector methods. Comput. Electron. Agric. 27, 401–406.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174. http://dx.doi.org/10.2307/2529310.

Moisen, G.G., Frescino, T.S., 2002. Comparing five modelling techniques for predicting forest characteristics. Ecol. Modell. 157, 209–225. http://dx.doi.org/10.1016/S0304-3800(02)00197-7.

Phillips, P.D., Brash, T.E., Yasman, I., Subagyo, P., van Gardingen, P.R., 2003. An individual-based spatially explicit tree growth model for forests in East Kalimantan (Indonesian Borneo). Ecol. Modell. 159, 1–26. http://dx.doi.org/10.1016/S0304-3800(02)00126-6.

Phillips, P.D., De Azevedo, C.P., Degen, B., Thompson, I.S., Silva, J.N.M., Van Gardingen, P.R., 2004. An individual-based spatially explicit simulation model for strategic forest management planning in the eastern Amazon. Ecol. Modell. 173, 335–354. http://dx.doi.org/10.1016/j.ecolmodel.2003.09.023.

Reis, L.P., Ruschel, A.R., Coelho, A.A., Luz, A.S., Da Martins-da-Silva, R.C.V., 2010. Avaliação do potencial madeireiro na Floresta Nacional do Tapajós após 28 anos da exploração florestal. Pesqui. Florest. Bras. 30, 265–281. http://dx.doi.org/10.4336/2010.pfb.30.64.265.

Reis, L.P., dos Reis, P.C.M., Ruschel, A.R., Silva, J.N.M., de Carvalho, J.O.P., de Souza, A.L., Soares, M.H.M., Miyahara, R.K.N., 2015. Forest dynamics in the eastern amazon with special reference to sapotaceae species. Floresta 45, 567–576. http://dx.doi.org/10.5380/rf.v45i3.35947.

Reis, L.P., de Souza, A.L., Mazzei, L., dos Reis, P.C.M., Leite, H.G., Soares, C.P.B., Torres, C.M.M.E., da Silva, L.F., Ruschel, A.R., 2016. Prognosis on the diameter of individual trees on the eastern region of the amazon using artificial neural networks. For. Ecol. Manage. 382, 161–167. http://dx.doi.org/10.1016/j.foreco.2016.10.022.

Richards, M., McDonald, A.J.S., Aitkenhead, M.J., 2008. Optimisation of competition indices using simulated annealing and artificial neural networks. Ecol. Modell. 214, 375–384. http://dx.doi.org/10.1016/j.ecolmodel.2008.03.008.

Rubin, B.D., Manion, P.D., Faber-Langendoen, D., 2006. Diameter distributions and structural sustainability in forests. For. Ecol. Manage. http://dx.doi.org/10.1016/j.foreco.2005.10.049.

Silva, J.N.M., Lopes, J.C.A., Oliveira, L.C., Silva, S.M.A., Carvalho, J.O.P., Costa, D.H.M., Melo, M.S., Tavares, M.J.M., 2005. Diretrizes Simplificadas para Instalação e Medição de Parcelas Permanentes em Florestas Naturais da Amazônia Brasileira, Embrapa/ITTO.

Silva, I.N., Spatti, D.H., Flauzino, R.A., 2010. Redes Neurais Artificiais: Para Engenharia e Ciências Aplicadas. ARTLIBER, São Paulo.

Stage, A., 1973. Prognosis Model for Stand Development, USDA for Serv. Res. Pap., Washington, DC. http://dx.doi.org/10.5962/bhl.title.69018. (INT-137).

Stan, A.B., Daniels, L.D., 2014. Growth releases across a natural canopy gap-forest gradient in old-growth forests. For. Ecol. Manage. 313, 98–103. http://dx.doi.org/10.1016/j.foreco.2013.11.004.

StatSoft Inc, 2016. StatSoft Stat. Data Anal. Softw. Syst. pp. 13. ([WWW Document]. URL). http://www.statsoft.com/.

Valle, D., Phillips, P., Vidal, E., Schulze, M., Grogan, J., Sales, M., van Gardingen, P., 2007. Adaptation of a spatially explicit individual tree-based growth and yield model and long-term comparison between reduced-impact and conventional logging in eastern Amazonia. Brazil. For. Ecol. Manage. 243, 187–198. http://dx.doi.org/10.1016/j.foreco.2007.02.023.

Vanclay, J.K., 1994. Modelling forest growth and yield: applications to mixed tropical forests. Modell. For. Growth Yield Appl. Mixed Trop. For. 312.

Wernsdörfer, H., Rossi, V., Cornu, G., Oddou-Muratorio, S., Gourlet-Fleury, S., 2008. Impact of uncertainty in tree mortality on the predictions of a tropical forest dynamics model. Ecol. Modell. 218, 290–306. http://dx.doi.org/10.1016/j.ecolmodel.2008.07.017.