Revista
ÁRVORE
ISSN 1806 - 9088

Scientific Article

ENHANCING SITE QUALITY CLASSIFICATION IN EUCALYPTUS PLANTATIONS: A MULTIVARIATE APPROACH USING SELF-ORGANIZING MAPS

Eulália Aparecida Silva[2], Pedro Alves Marques[3*], Luiana Rolim de Azevedo[3], Geovanna Sales Pereira Rabelo[4], Gilciano Saraiva Nogueira[5], Helio Garcia Leite[5] and Marcio Leles Romarco de Oliveira[6]

1 Received on 13.11.2025 accepted for publication on 14.03.2026. Editors: Carlos Moreira Miquelino Eleto Torres and Bruno Leão Said Schettini.
2 Aperam BioEnergia, Capelinha, Minas Gerais, Brasil. E-mail: <eulalia.a.silva@gmail.com>.
3 Universidade Federal de Viçosa, Programa de Pós-Graduação em Ciência Florestal, Viçosa, Minas Gerais, Brasil. E-mail: <pedro.alves.marques@ufv.br> and <luiana.azevedo@ufv.br>.
4 Universidade Federal dos Vales do Jequitinhonha e Mucuri, Graduando em Engenharia Florestal, Diamantina, Minas Gerais, Brasil. E-mail: <geovanna.rabelo@ufvjm.edu.br>.
5 Universidade Federal de Viçosa, Departamento de Engenharia Florestal, Viçosa, Minas Gerais, Brasil. E-mail: <gilciano.nogueira@ufv.br> and <hgleite@ufv.br>.
6 Universidade Federal dos Vales do Jequitinhonha e Mucuri, Departamento de Engenharia Florestal, Diamantina, Minas Gerais, Brasil. E-mail: <marcioromarco@ufvjm.edu.br>.
*Corresponding author.

ABSTRACT

The classification of productive capacity is a fundamental component of forest management, but traditional univariate methods, such as the guide-curve, may oversimplify the complex interactions of factors determining site productivity. Therefore, this study addresses a research gap by assessing the potential of Kohonen Neural Networks (SOM), a multivariate artificial intelligence technique, as an alternative for stratifying commercial eucalyptus stands. The objective was to compare the classification obtained by the guide-curve method with the artificial neural network (ANN) approach, assessing the superiority of the latter. To this end, data from 750 eucalyptus stands in Minas Gerais, Brazil, were classified using both the guide-curve method (Schumacher's model) and an ANN (SOM) with subsequent hierarchical clustering. Validation and comparison were performed using discriminant analysis and a contingency matrix. The results indicated that the ANN provided a more accurate stratification, notably by reclassifying 26% of the stands from the guide-curve's Medium class to the Low class and refining the High-productivity class into a more elite group. The cohesion of the ANN-derived clusters was validated by discriminant analysis, which achieved an overall accuracy of 69.9%. It is concluded that the ANN is superior to the traditional method, providing a more realistic, multidimensional classification with greater sensitivity in detecting low-productivity areas and greater specificity in identifying elite stands. The methodology is thus established as a strategic tool for enhancing planning and optimizing operations in precision silviculture.

Keywords: Site quality; Kohonen network; Precision silviculture

How to cite:
Silva, E. A., Marques, P. A., Azevedo, L. R. de, Rabelo, G. S. P., Nogueira, G. S., Leite, H. G., & Oliveira, M. L. R. de. (2026). Enhancing site quality classification in eucalyptus plantations: A multivariate approach using self-organizing maps. *Revista Árvore, 50*(1). https://doi.org/10.53661/1806-9088202650264038

APRIMORANDO A CLASSIFICAÇÃO DE SÍTIOS EM PLANTIOS DE EUCALIPTO: UMA ABORDAGEM MULTIVARIADA COM MAPAS AUTO-ORGANIZÁVEIS

RESUMO A classificação da capacidade produtiva é um pilar para o manejo florestal, mas métodos tradicionais univariados, como a curva-guia, podem simplificar a complexa interação de fatores que definem a produtividade de um sítio. Diante disso, este estudo preenche uma lacuna ao avaliar o potencial das Redes Neurais de Kohonen (SOM), uma técnica de inteligência artificial multivariada, como alternativa para a estratificação de povoamentos comerciais de eucalipto. O objetivo foi comparar a classificação obtida pelo método da curva-guia com a abordagem via rede neural artificial (RNA), avaliando a superioridade desta última. Para isso, dados de 750 talhões de eucalipto em Minas Gerais foram classificados pelo método da curva-guia (modelo de Schumacher) e por uma RNA do tipo SOM, com posterior agrupamento hierárquico. A validação e a comparação foram realizadas por análise discriminante e matriz de contingência. Os resultados indicaram que a RNA promoveu uma estratificação mais acurada, reclassificando 26% dos talhões da classe Média (curva-guia) para a classe Inferior e refinando a classe Superior para um grupo mais seleto. A coesão dos clusters da RNA foi validada por análise discriminante, que alcançou uma acurácia geral de 69,9%. Conclui-se que a RNA é superior ao método tradicional, fornecendo uma classificação multidimensional mais realista, com maior sensibilidade na detecção de áreas de baixa produtividade e maior especificidade na identificação de talhões de elite. A metodologia estabelece-se, portanto, como uma ferramenta estratégica para o aprimoramento do planejamento e a otimização de operações na silvicultura de precisão.

Palavras-Chave: Qualidade de local; Rede de Kohonen; Silvicultura de precisão

1. INTRODUCTION

Sustainable forest management requires the assessment of a series of conditions to support decision-making, from planning to the implementation of silvicultural prescriptions (Skovsgaard & Vanclay, 2013). Given the increasing demand for accurate assessments of forest productivity, one of the important indicators in this process is productive capacity, which represents the potential of a site to produce timber or other forest products (Campos & Leite, 2025; Gopalakrishnan et al., 2019). The classification of areas according to their productive potential is, therefore, a fundamental step in management, serving as a basis for short-, medium-, and long-term planning of forest activities (Molina-Valero et al., 2019; Scolforo, 1992).

Historically, the most common method in forest plantations for this classification has been the site index, defined as the average dominant height of a stand at a reference age (Campos & Leite, 2025; Weiskittel et al., 2011). Dominant height is preferred as an indicator of site quality because, within certain limits, it is not influenced by stand density or silvicultural interventions (García, 1995) and presents a strong correlation with volume production (Clutter et al., 1983). However, forest productivity modeling has evolved from these traditional age-height relationships to sophisticated techniques that employ remote sensing and machine learning, such as Airborne Laser Scanning (ALS) and neural networks, which allow for large-scale, spatially explicit productivity mapping (Mensah et al., 2023; Kanga, 2023; Rizzo-Martín et al., 2023).

Despite these advances, reliance on a single variable to represent a complex system has disadvantages. Height measurements can be imprecise (Leite et al., 2011; Binoti et al., 2013), and the premise that all factors affecting growth can be represented by a single variable is a simplification. Recent studies highlight persistent challenges in integrating complex environmental variables and remote sensing data to improve site index predictions (Fiandino et al., 2020; Huang et al., 2024). Consequently, a research gap exists concerning the effective use of Self-Organizing Maps (SOM) for site index modeling, particularly regarding their capacity to handle non-linear relationships and spatial heterogeneity (Sun et al., 2022).

Artificial Intelligence tools, such as Artificial Neural Networks (ANNs), emerge as a promising alternative, as they can simultaneously process multiple quantitative and qualitative variables (Cosenza et al., 2015). Among ANN architectures, Kohonen Self-Organizing Maps, or SOMs, are an unsupervised neural network technique suitable for data clustering and pattern recognition tasks (Costa et al., 2011; Sun et al., 2022). Although machine learning methods such as Random Forest and Boosted Regression Trees have shown promise, the comparative advantages and limitations of SOMs in this context remain underexplored (Yang & Meng, 2022; Duan et al., 2022). The lack of consensus and empirical evaluations of SOM-based models limits their wider adoption, justifying investigation into the applicability of this tool (Santos et al., 2025; Penner et al., 2023).

Therefore, the objective of this study is to compare the classification of productive capacity in *Eucalyptus* stands using the traditional guide-curve method with a Kohonen Self-Organizing Map (SOM) approach and evaluate the potential of the latter as a tool for forest site stratification. While the present model relies solely on dendrometric inventory variables, this approach aims to establish a robust analytical base that enables, as a potential future refinement, the synergistic integration of environmental variables (edaphic and climatic) and remote sensing data.

2. MATERIAL AND METHODS

2.1 Study area and database

The study was conducted using data from continuous forest inventories of Eucalyptus ssp. clones in commercial plantations located in Minas Gerais, Brazil. The region encompasses diverse site conditions representative of typical eucalyptus growing areas in southeastern Brazil, characterized by a tropical highland climate (Aw/Cwa/Cwb) with well-defined dry and wet seasons (October to March) (Alvares et al., 2013). The study area spans elevations ranging from 600 to 1,200 meters, with mean annual precipitation between 1,200 and 1,800 mm and mean annual temperatures from 18°C to 24°C (Alvares et al., 2013). Soils are predominantly Oxisols (Latossolos) and Inceptisols (Cambissolos), with varying fertility levels and physical properties (Amaral et al., 2004) that contribute to the observed productivity differences across the landscape (Gonçalves et al., 2004).

The database for this study originated from 2,334 permanent plots distributed across 1,703 stands. For the final analysis, a dataset from the last measurement cycle was used, comprising 750 stands covering a total area of 17,764.14 hectares. This dataset included the following stand-level variables: age (A, months), dominant height (h_d, m), total height (h_t, m), quadratic mean diameter (q, cm, derived from DBH), basal area (G, $m^2 \cdot ha^{-1}$), total volume with bark (V, $m^3 \cdot ha^{-1}$), and the number of trees per hectare (N).

2.2 Site classification by the Guide-Curve method (GC)

Traditional site classification was based on the guide-curve method, using the Schumacher (1939) model fitted to dominant height (h_d) and age (in months) data. The model was fitted in its linearized form via ordinary least squares (OLS) (Eq. 1). The goodness-of-fit was evaluated by the coefficient of determination (R^2).

$$ln(h_d) = \beta_0 + \beta_1(A^{-1}) + \varepsilon \qquad \text{(Eq. 1)}$$

To ensure the most representative fit for each sampling unit, a hierarchical strategy was adopted. Priority was given to fitting the model at the stand level, provided that a minimum of three longitudinal measurements were available. Otherwise, the fit was sought sequentially at progressively broader stratification levels.

The site index (S) was calculated for an index age of 72 months (6 years), and the family of anamorphic site index curves was generated based on Equation 2.

$$h_d = \exp\left[ln(S) + \widehat{\beta_1}\left(A^{-1} - {A_i}^{-1}\right)\right] \qquad \text{(Eq. 2)}$$

Where: ln is the natural logarithm; h_d is the dominant height (m); A is the age (months); β_0 and β_1 are the model parameters; S is the site index at the index age A_i ; and ε is the random error.

Subsequently, the 750 stands were stratified into three productivity classes (High, Medium, and Low) by dividing the total range of the observed dominant height into three equal intervals.

2.3 Site classification via Artificial Neural Network (ANN)

An unsupervised Artificial Neural Network, specifically a Kohonen Self-Organizing Map (SOM), was used to classify the stands based on their multidimensional dendrometric profile. An unsupervised Artificial Neural Network, specifically a Kohonen Self-Organizing Map (SOM), was used to classify the stands based on their multidimensional dendrometric profile.

2.3.1 *Data pre-processing and validation*

The six main dendrometric variables (V, G, h_d, h_t, q, N) were used as inputs for the network. The data were standardized to have zero mean and unit variance to ensure that all variables contributed equally to the training process, regardless of their original scale. To select the optimal model and prevent overfitting, the dataset was randomly partitioned into a training set (80% of the data) and a validation set (20%).

2.3.2 *SOM architecture and training*

The SOM network was configured with a total of 132 neurons (an 11x12 grid), a size determined following the $5\sqrt{n}$ heuristic, where n is the number of input samples (Vesanto et al., 2000). These neurons were arranged in a hexagonal topology, a structure that provides more uniform distances between adjacent neurons compared to a rectangular grid, thus being preferential for exploratory data analysis (Kohonen, 2001). The network was trained for 1000 epochs using a batch algorithm. The learning rate decayed linearly from an initial value of 0.05 to a final value of 0.01. Model performance was monitored at each epoch by calculating the Mean Quantization Error (MQE) on the validation set. The final model was selected as the one corresponding to the epoch with the minimum MQE on the validation set, ensuring optimal generalization capacity.

2.3.3 *Clustering and definition of productive classes*

After training, the SOM weight vectors, which represent the prototypical stand profiles learned by the network, were clustered using a hierarchical clustering algorithm (Ward's method, Euclidean distance). The dendrogram was cut to define three distinct clusters, ensuring comparability with the three classes generated by the traditional guide-curve method. These three clusters were then interpreted and labeled as High, Medium, and Low productivity classes.

2.4 Comparative analysis and statistical validation

To validate and compare the two classification methodologies, two statistical analyses were performed.

First, a Linear Discriminant Analysis (LDA) was used to evaluate the statistical division of the three clusters generated by the ANN. The ANN-derived classes were used as the response variable, and the original dendrometric variables were used as predictors. Model performance was evaluated using leave-one-out cross-validation, with overall accuracy and the confusion matrix serving as metrics for cluster cohesion.

Second, a contingency matrix (cross-tabulation table) was constructed to directly compare the stand-by-stand allocation between the guide-curve method and the ANN method. This matrix was used to quantify the overall agreement and to identify systematic patterns of disagreement and reclassification between the two approaches.

2.5 Software and analysis

All data processing, modeling, and statistical analyses were conducted in the R software environment (v4.4.1) (R Core Team, 2024). The Self-Organizing Map was trained and analyzed using the kohonen package (Wehrens & Kruisselbrink, 2018), and the discriminant analysis was performed with the MASS package (Venables & Ripley, 2002).

3. RESULTS

3.1 Site classification by the Guide-Curve method (GC)

The relationship between dominant height (h_d) and age (A) was successfully modeled using the Schumacher equation, which demonstrated a high goodness-of-fit, with a mean coefficient of determination (R^2) of 0.93. The mean estimated parameters for the model were $\beta0 = 3.715219$ and $\beta1 = -29.236196$, resulting in the final equation used for the construction of the site index curves (Equation 3; Figure 1):

$$ln(h_d) = ln(S) - 29{,}236196(A^{-1} - 72^{-1}) \quad \text{(Eq. 3)}$$

Based on the adjusted model, the 750 stands were stratified into three productive capacity classes, defined by the total range of dominant height at the index age, resulting in an interval of 4.8 m for each class. The stratification revealed a productive structure with a high concentration in the intermediate class. Class I (High), with h_d ranging from 30.3 to 35.0 m, represented 11% of the stands

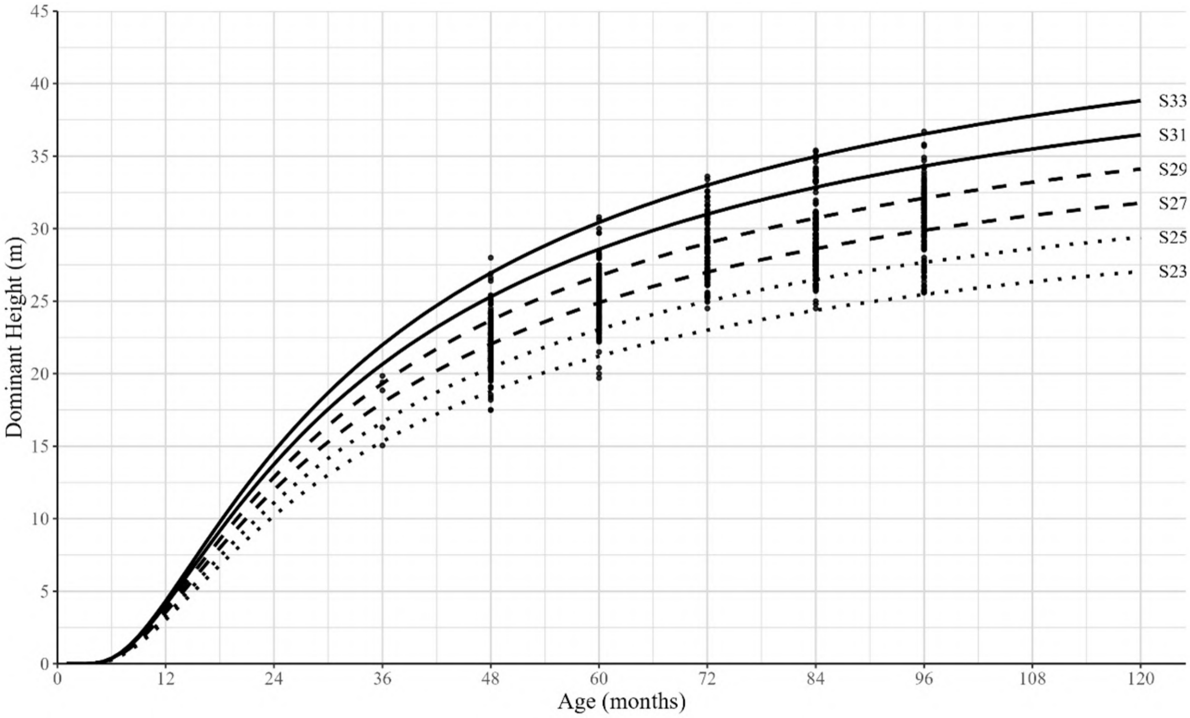


Figure 1. Site index curves for *Eucalyptus* ssp. obtained by the guide-curve method for a reference age of 72 months
Figura 1. Curvas de índices de local para *Eucalyptus* ssp. obtidas pelo método da curva-guia para uma idade índice de 72 meses

and 11% of the total area. Class II (Medium), with h_d between 25.5 and 30.2 m, was the most representative, encompassing 68% of the stands and 67% of the area.

Finally, Class III (Low), with h_d from 20.7 to 25.4 m, accounted for 21% of the stands and 21% of the area.

The dendrometric characterization of each class at the last measurement is detailed in Table 1. As expected, Class I exhibited the highest mean values for all production variables, reaching a volume (V) of 322.72 $m^3 \cdot ha^{-1}$ at 96 months. In contrast, Class III recorded the lowest values, with a mean volume of 189.87 $m^3 \cdot ha^{-1}$ at the same age, while Class II showed intermediate values.

3.2 Neural Network training and validation

The SOM training process, as described in the methodology, was monitored to evaluate the model's performance (Figure 2). The learning curve for the training set (Figure 2, blue line) demonstrated the expected monotonic decline in the Mean Quantization Error (MQE), with a steep drop in the initial iterations, indicating a rapid topological organization phase of the network. The error continued to decrease more gradually, stabilizing at a plateau of approximately 0.32 at the end of the process.

The error curve for the validation set (Figure 2, red line) exhibited a convergent

Table 1. Description of the productive capacity classes (high-I, medium-II, and low-III) obtained by the guide-curve method, with number of stands (Stands), area, and mean values of dendrometric variables by age at the last measurement
Tabela 1. Descrição das classes de capacidade produtiva (superior-I, média-II e inferior-III) obtidas pelo método da curva-guia, com número de talhões (Stands), área e valores médios das variáveis dendrométricas por idade na última medição

Class	Age (months)	Area (ha)	Stands	V ($m^3.ha^{-1}$)	B ($m^3.ha^{-1}$)	Ht (m)	Hd (m)	q (cm)
I	36	56.80	3	88.14	10.83	17.6	19.4	11.4
	48	223.28	9	172.40	14.86	22.3	25.7	13.2
	60	327.45	13	234.32	18.08	24.6	28.9	14.1
	72	641.13	24	346.62	21.03	28.8	31.6	16.0
	84	601.97	25	327.22	23.39	30.7	33.8	16.5
	96	178.25	11	322.72	22.31	29.9	34.7	16.4
	Subtotal	2,028.88	85					
II	36	31.13	2	54.03	8.54	14.4	15.7	10.2
	48	3,352.04	125	132.72	13.29	20.5	22.5	12.2
	60	2,828.51	104	187.39	15.93	23.4	25.6	13.3
	72	1,544.23	67	232.51	18.41	25.6	27.7	14.3
	84	1,874.19	94	222.50	18.61	25.0	29.2	15.1
	96	2,344.19	119	252.09	19.71	27.2	31.1	15.4
	Subtotal	11,974.29	511					
III	48	1,543.60	54	104.91	11.92	19.0	20.5	11.7
	60	1,066.95	39	147.03	14.27	21.3	22.9	12.6
	72	142.36	7	186.53	15.98	23.0	25.1	13.5
	84	467.96	29	174.04	16.61	22.1	26.5	14.4
	96	540.10	25	189.87	17.95	23.1	26.6	14.9
	Subtotal	3,760.97	154					
Total		17,764.14	750					

behavior. After an initial drop similar to that of the training set, the validation error stabilized at a higher level, around 0.60, starting from approximately 300 iterations. The optimal performance point was precisely identified at iteration 688, where the MQE for the validation set reached its minimum value of 0.539. The absence of an upward trend in the validation error after this point, even with the continuation of training, is quantitative evidence of the absence of overfitting.

Based on this result, the model corresponding to iteration 688 was selected as the final model for the subsequent stages of classification and data structure analysis.

3.3 Analysis of map topology and data structure

Following model validation, the structure of the trained neural map was inspected through a set of three topological diagnostic plots (Figure 3). The hit map (Figure 3-A), which illustrates the frequency with which each neuron was selected as the best matching unit (BMU) for the data samples, revealed a heterogeneous distribution. The presence of multiple neurons with high activation frequency (shades of yellow and white), distributed in different regions of the map, indicates the existence of prototypes or recurrent dendrometric profiles in the dataset, suggesting a natural clustering structure.

In contrast, the U-Matrix (Figure 3-B), which visualizes the mean Euclidean distance between adjacent neurons, presented a homogeneous topography with high values (shades of red). This result indicates that inter-neuronal distances are consistently high across the grid, without the formation of "valleys" (light-colored areas) that would characterize the presence of well-defined clusters clearly separated by "ridges" (dark-colored areas).

This finding was corroborated by the mapping quality map (Figure 3-C), which represents the mean intra-neuronal

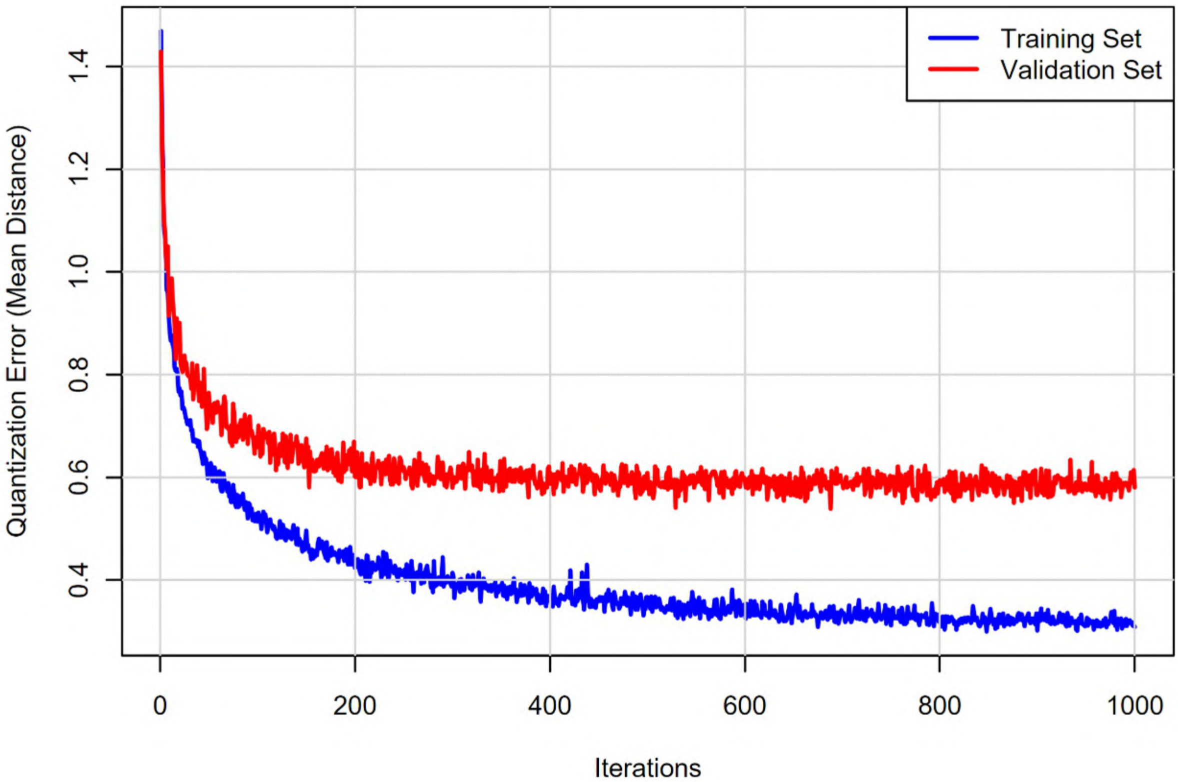


Figure 2. Evolution of the mean quantization error (MQE) during the Artificial Neural Network training for the training (80%) and validation (20%) datasets. Iteration 688, which showed the minimum MQE on the validation set, was selected for the final model
Figura 2. Evolução do erro de quantização médio (EQM) durante o treinamento da Rede Neural Artificial para os conjuntos de dados de treinamento (80%) e validação (20%). A iteração 688, que apresentou o menor EQM na validação, foi selecionada para o modelo final

quantization error. The predominance of dark colors in this map signals a relatively high mean error for most neurons, indicating considerable dispersion of the data samples around their respective weight vectors. In summary, the topological results suggest a data structure that, although presenting regions of higher density, is characterized by high internal variance and a gradual transition between profiles, rather than discrete, low-variance clusters.

3.4 Analysis of map structure and network clustering

The analysis of the internal structure of the trained neural map was performed using component planes (Figure 4), which visualize the gradient of each dendrometric variable within the network topology. The production and dimension variables – Volume (V), Basal Area (G), Heights (h_t e h_d), and Quadratic Mean Diameter (q – exhibited a similar and strongly correlated distribution pattern, with high values (shades of red) concentrated in the upper-left corner of the map and low values (shades of blue) in the right portion. In contrast, the density variable (N) showed an opposite gradient, with the highest values (red) on the right side of the map, coinciding with the areas of lower individual growth.

This clear spatial organization of the data into gradients formed the basis for the subsequent segmentation of the network into homogeneous clusters. The application of a hierarchical clustering algorithm on the network weight vectors resulted in the identification of three distinct clusters (Figure 5), which were labeled according to their productive profile: Medium Productivity Class (Cluster 1), Low Productivity Class (Cluster 2), and High Productivity Class (Cluster 3).

The spatial distribution of the clusters (Figure 5) showed that the Medium

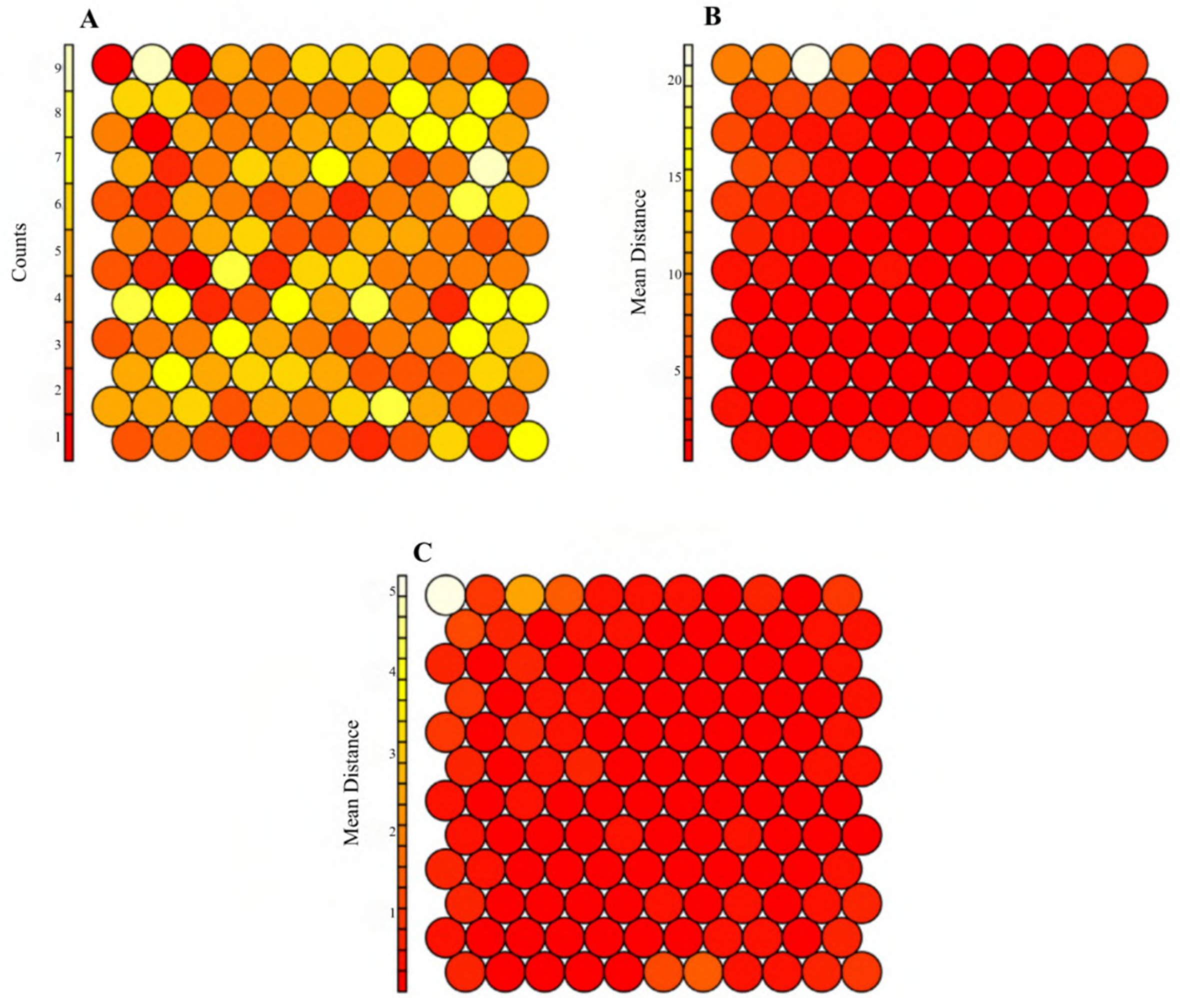
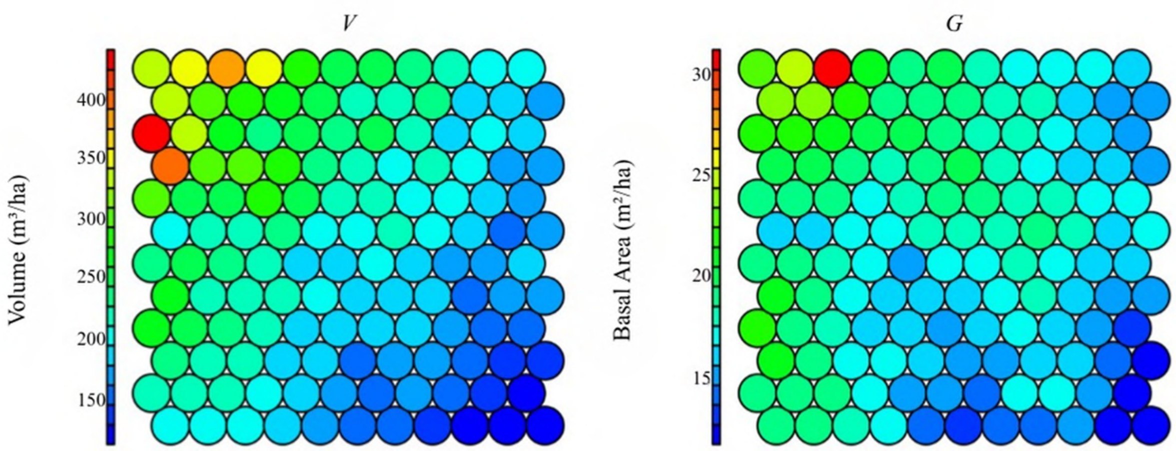


Figure 3. Topological diagnostic maps of the trained SOM network: (A) Hit map, illustrating the stand mapping frequency; (B) U-Matrix, representing the mean distance between neighboring neurons; and (C) Quality map, representing the mean intra-neuronal quantization error

Figura 3. Mapas de diagnóstico topológico da rede SOM treinada: (A) Mapa de contagem de ativações (hit map), ilustrando a frequência de mapeamento dos talhões; (B) Matriz-U, representando a distância média entre neurônios vizinhos; e (C) Mapa de qualidade, representando o erro de quantização médio intra-neuronal



Cont...

...Cont

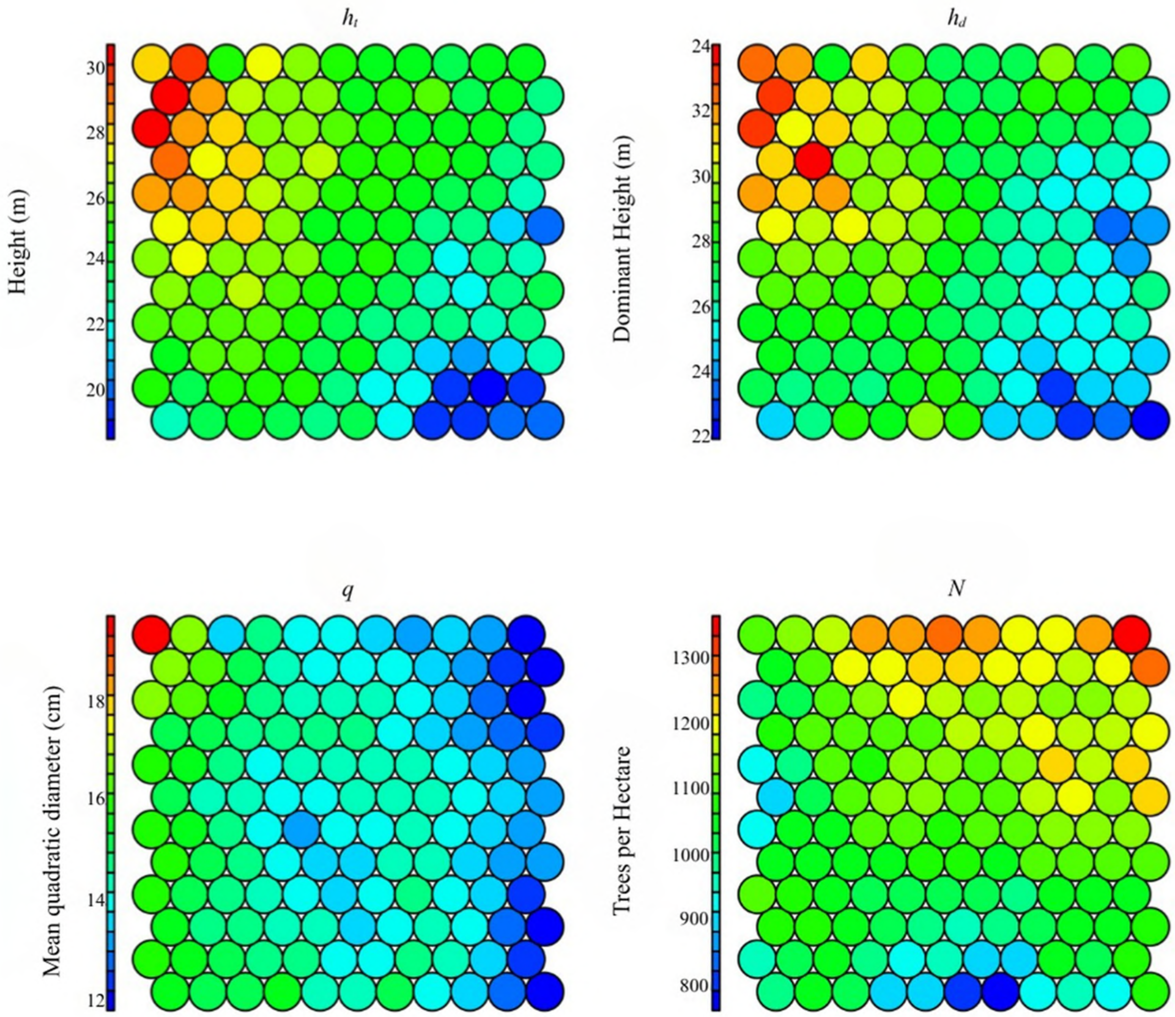


Figure 4. Component planes of the trained SOM network for the six dendrometric variables. The color gradient represents the variation of each neuron's synaptic weight values, from the lowest (blue) to the highest (red)

Figura 4. Planos de componentes da rede SOM treinada para as seis variáveis dendrométricas. O gradiente de cores representa a variação dos valores dos pesos sinápticos de cada neurônio, do menor (azul) ao maior (vermelho)

Productivity Class (green) constituted the largest grouping, occupying the central and lower regions. The Low Productivity Class (orange) was mapped to the right portion, while the High Productivity Class (purple) formed a cohesive group in the upper-left corner. The silvicultural characterization of each class was confirmed by the analysis of the mean variable profiles (Figure 4). The High Productivity Class presented prominent segments for the dimension and volume variables. The Low Productivity Class, in turn, was dominated by the segment of the N variable, with low values for the others. The Medium Productivity Class exhibited an intermediate dendrometric profile for all variables.

3.5 Site classification by the Artificial Neural Network (ANN)

The application of the trained and validated Self-Organizing Map (SOM) model allowed the classification of the 750 stands into three productive capacity classes, according to the data structure identified by the network itself. Table 2 presents the stand distribution and the dendrometric characterization of each class at the last measurement.

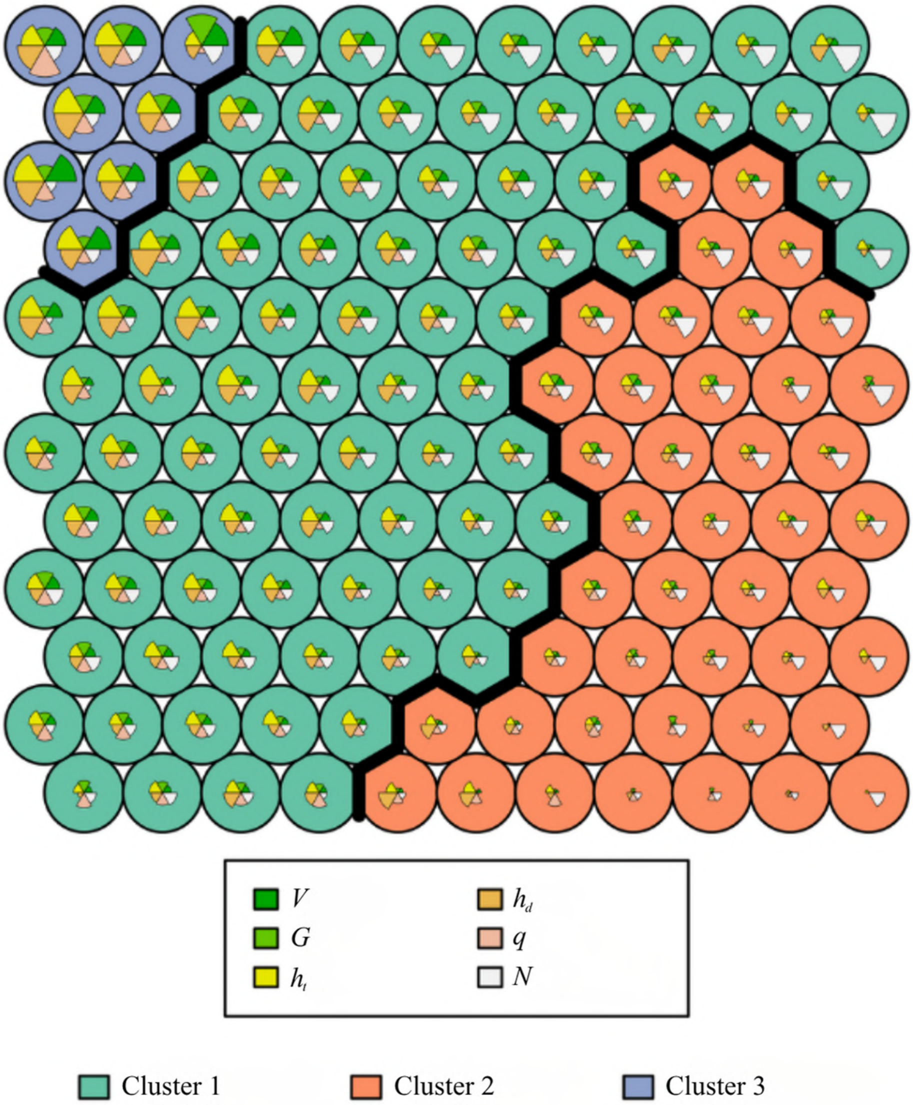


Figure 5. Result of the hierarchical clustering of the SOM network's neurons: Cluster map, with the boundaries between the High (3), Medium (1), and Low (2) productivity classes; and Dendrometric variable profiles within each cluster, represented by segment plots
Figura 5. Resultado do agrupamento hierárquico dos neurônios da rede SOM: Mapa de clusters, com as fronteiras entre as classes de produtividade Alta (3), Média (1) e Baixa (2); e Perfis das variáveis dendrométricas dentro de cada cluster, representados por gráficos de segmento

The ANN generated a stratification distinct from that obtained by the guide-curve method. Class I (High), as defined by the ANN, grouped 39 stands (5% of the total), representing the smallest and most select class. Class II (Medium) remained the most populous, with 452 stands (60%), while Class III (Low) was composed of 259 stands (35%).

The productive profiles of each class (Table 2) are distinct. At 84 months, Class I recorded the highest mean volume (V) of 358.82 $m^3 \cdot ha^{-1}$. At the same age, Class II presented a volume of 238.50 $m^3 \cdot ha^{-1}$, and Class III reached 189.61 $m^3 \cdot ha^{-1}$.

3.6 Validation of the ANN classification by discriminant analysis

The Linear Discriminant Analysis (LDA) with leave-one-out cross-validation, performed to test whether the original dendrometric variables could correctly predict a stand's class as defined by the ANN, resulted in an overall classification accuracy of 69.9%. The analysis of the confusion matrix (Table 3) revealed that Class II (Medium) was the most cohesive and statistically separable group, with 384 out of 452 stands (85%) correctly classified by the discriminant model.

Table 2. Description of productive capacity classes (high-I, medium-II, and low-III) obtained by the artificial neural network (ANN), with number of stands (Stands), area, and mean values of dendrometric variables by age at the last measurement

Tabela 2. Descrição das classes de capacidade produtiva (superior-I, média-II e inferior-III) obtidas por meio de rede neural artificial (RNA), com número de talhões (Stands), área e valores médios das variáveis dendrométricas por idade na última medição

Class	Age (months)	Area (ha)	Stands	V ($m^3.ha^{-1}$)	B ($m^3.ha^{-1}$)	Ht (m)	Hd (m)	q (cm)
I	36	56.8	3	88.14	10.83	17.6	19.4	11.4
	48	44.04	2	225.74	22.06	21.4	23.0	13.2
	60	10.42	1	264.27	21.71	26.0	30.0	15.4
	72	351.27	16	382.28	21.79	29.5	31.8	16.6
	84	352.35	16	358.82	25.15	31.4	33.8	17.0
	Subtotal	822.05	39					
II	36	31.13	2	54.03	8.54	14.4	15.7	10.2
	48	3,440.87	123	134.90	13.34	20.7	22.7	12.3
	60	2,969.68	108	194.59	16.23	23.6	26.0	13.4
	72	1,659.33	68	242.91	18.74	26.0	28.3	14.5
	84	1,499.74	69	238.50	19.52	26.3	29.8	15.5
	96	1,584.31	82	272.79	20.42	28.1	31.9	15.7
	Subtotal	11,185.06	452					
III	48	1,634.01	63	107.35	11.96	19.1	20.7	11.7
	60	1,242.81	47	148.69	14.34	21.4	23.2	12.7
	72	317.12	14	183.46	16.22	23.5	25.8	13.7
	84	1,092.03	63	189.61	16.93	22.9	27.9	14.5
	96	1,471.06	72	215.59	18.57	25.1	29.1	14.9
	Subtotal	5,757.03	259					
Total		17,764.14	750					

The extreme classes, Class I (High) and Class III (Low), showed considerable overlap with the intermediate class, presenting classification accuracies of 49% and 47%, respectively. Notably, there was no confusion between the extreme classes, as no sample from Class III was erroneously classified as Class I.

3.7 Concordance analysis between the Guide-Curve (GC) and ANN classifications

The final stage of the analysis consisted of a direct comparison of the allocation of each of the 750 stands between the guide-curve (GC) method and the artificial neural network (ANN). The contingency matrix (Table 4), which cross-tabulates the classification of both methods, summarizes the agreement and disagreement between the two approaches.

An overall agreement of 71.9% (539 stands) was observed between the two

Table 3. Confusion matrix from the Linear Discriminant Analysis (LDA) for the validation of the ANN clusters. Rows represent the actual class (defined by the ANN) and columns represent the class predicted by the LDA model. C (%) represents the concordance (accuracy) for each class

Tabela 3. Matriz de confusão da Análise Discriminante Linear (LDA) para a validação dos clusters da RNA. As linhas representam a classe real (definida pela RNA) e as colunas representam a classe prevista pelo modelo LDA. C (%) representa a concordância (acurácia) para cada classe

Class	I	II	III	Total	C (%)
I	19	15	5	85	48.7
II	9	384	59	511	85.0
III	0	138	121	154	46.7
Total	28	537	155	750	69.9

methods. Agreement was most pronounced in the Medium Class (376 stands) and the Low Class (126 stands). The analysis of disagreements, however, revealed a systematic reclassification pattern by the ANN.

The main divergence occurred in the reclassification of stands originally classified as Class II (Medium) by the guide-curve (GC) method. From this stratum, 133 stands (26%) were reclassified as Class III (Low) by the ANN. Another divergence was observed in Class I (High): of the 85 stands considered high-productivity by the GC method, the ANN reclassified 48 stands (56%) as Class II (Medium), keeping only 37 stands in the most productive class.

Despite reclassifications between adjacent classes, there was total agreement in the separation of productive extremes: no stand classified as Class I (High) by the GC was classified as Class III (Low) by the ANN, and vice versa.

Table 4. Contingency matrix comparing the allocation of the 750 stands between the Guide-Curve method (GC, rows) and the Artificial Neural Network method (ANN, columns). C (%) represents the classification agreement percentage for each Guide-Curve method class
Tabela 4. Matriz de contingência comparando a alocação dos 750 talhões entre o método da Curva-Guia (CG, linhas) e o da Rede Neural Artificial (RNA, colunas). C (%) representa a porcentagem de concordância de classificação para cada classe do método da Curva-Guia

Class	I	II	III	Total	C (%)
I	37	48	0	85	42.35
II	2	376	133	511	77.50
III	0	28	126	154	24.03
Total	39	452	259	750	71.86

4. DISCUSSION

4.1 Forest stand structure from the perspective of the Guide-Curve method

Site classification using the guide-curve method, based on the Schumacher model, provided a statistically valid stratification of the height growth potential of the eucalyptus stands. The high mean value of the coefficient of determination (R^2 = 0.93) confirms the strong relationship between dominant height and age, validating the use of this method as a standard tool for the productive zoning of the area (Oliveira et al., 2008; Campos & Leite, 2025).

The concentration of most stands in the medium class (Class II) by the guide-curve method represents an expected pattern, which can be associated with the central limit theorem. Since forest productivity is influenced by multiple factors (e.g., genetics, soil, microclimate), their combined effect tends towards a normal distribution, with most observations clustering around the mean (Gujarati & Porter, 2011). The extreme high and low performance classes (I and III), therefore, naturally comprise a smaller fraction of the productive landscape.

It is essential to emphasize, however, the inherent limitations of the traditional guide-curve method employed. First, it is, by definition, univariate, assuming that dominant height, being little influenced by density, is a sufficient and integrated indicator of total site productivity. Second, the approach with the Schumacher model resulted in anamorphic site index curves, which assume that the shape of the growth curve is the same for all productivity classes, differing only by a proportional factor (Campos & Leite, 2025). Although these premises are widely accepted and used in forest management, they do not capture the variation in other structural dimensions of the stand, such as tree density, diameter distribution, and basal area, which are also crucial components of volumetric productivity (Gonçalves, 2022). The classification, therefore, exclusively reflects the potential for height growth, and not necessarily the complete multidimensional productive profile of the stand. This inherent limitation justifies the exploration of alternative methodologies, such as those based on artificial intelligence, which will be addressed subsequently.

4.2 Model robustness and generalization capacity

The analysis of the learning curves (Figure 2) is an indispensable step to attest to the validity and robustness of the proposed neural network model. The observed behavior validates the adequacy of the SOM's architecture and hyperparameters for modeling the structure of the dendrometric data, demonstrating that the model was capable of capturing generalizable patterns.

The stabilization of the validation error at a consistent plateau, without the characteristic increase of overfitting, is the main indicator that the model learned the structural "signal" present in the data rather than memorizing the random "noise" specific to the training subset (Prechelt, 2012). The use of a data partition for cross-validation confers statistical confidence to this conclusion, ensuring that the model's performance on unseen data is stable and predictable. This result is fundamental, as it legitimizes the neural map as a faithful representation of the sample universe (Leisch et al., 1998).

The selection of iteration 688 as the final model represents a methodological decision based on the well-known bias-variance trade-off (Yu et al., 2006). Although continuing the training could marginally reduce the error on the training set, the validation curve shows that such a gain would not translate into a better generalization capacity (Prechelt, 2012). The point of minimum error on validation is, therefore, the optimal point where the model maximizes its predictive performance on new samples (Hastie et al., 2009).

4.3 The continuous nature of productive variation in eucalyptus plantations

The joint analysis of the topological maps, when contextualized by the data source, offers a profound diagnosis not only of the data structure but also of the productive dynamics of these managed ecosystems. The apparent dichotomy between the identification of high-frequency prototypes in the hit map (Figure 3-A) and the absence of clear boundaries in the U-Matrix (Figure 3-B) solidifies a conclusion of great silvicultural relevance: the variation in productive capacity, even in a high-tech forest environment where factors like genetics (clones) and initial spacing are controlled, occurs along a continuous gradient.

In a forest production system, the emergence of a productive gradient rather than discrete classes points to the predominance of site edaphoclimatic variation (micro-variations in soil, relief, and water availability) and intra-stand competitive dynamics as the main drivers of productivity (Resende et al., 2016; Skovsgaard & Vanclay, 2013). The clusters indicated by the hit map are, from this perspective, the most common and stable profiles resulting from the interaction between genetic material and these local conditions—the phenotypes most frequently expressed in the landscape.

In this context, the homogeneity of high values in the U-Matrix (Figure 3-B) should not be interpreted as a failure in cluster separation, but rather as the faithful representation of gradual productivity transitions between neighboring stands. This finding is ecologically coherent with the continuous spatial variation of site factors. The SOM's ability to faithfully map this gradient structure without imposing artificial boundaries a priori constitutes one of its main advantages over clustering methods that assume spherical and well-separated groups (Giraudel & Lek, 2001; Gonçalves & Vigário, 2012; Costa, 2010).

The high quantization error (Figure 3-C) and the elevated inter-neuronal distances (Figure 3-B) reinforce that the model is representing a complex system. The high variance of the data, even within the densest regions, is the signature of the complex genotype-environment-management interaction.

4.4 Validation and interpretability of the ANN clusters

The overall accuracy of 69.9% obtained in the discriminant analysis, although moderate, is a significant result that validates the classification generated by the ANN. This value indicates that the clusters formed by the SOM are not random but represent a real and separable structure in the data, to the point that a linear model like LDA can predict which group a stand belongs to with nearly 70% accuracy.

The confusion matrix offers deeper insights. The high accuracy in classifying the High Productivity Class (Class I) indicates that this group, as defined by the ANN, has a very distinct and cohesive dendrometric profile, being easily separable from the others. This reinforces the conclusion that the ANN was effective in identifying a high-productivity group with exceptional growth characteristics.

The greater confusion between the Medium (Class II) and Low (Class III) classes is equally informative and consistent with the topological analysis of the map (U-Matrix). As previously discussed, forest data exist in a productive gradient. The overlap found by the LDA reflects the gradual transition zone between the medium and low productivity stands. A linear model like LDA has difficulty drawing an exact dividing line in this gradient, which justifies the use of a non-linear and topological technique like the SOM to find the primary structure. The ability of the ANN to extract three groups that are nonetheless recognizable by a classic statistical method attests to its effectiveness.

4.5 Superiority of the multivariate ANN classification over the univariate approach

The concordance analysis between the methodologies not only quantifies the differences but fundamentally elucidates the conceptual advantages of the multivariate and unsupervised ANN approach for the stratification of commercial forest sites. The overall agreement of 71.9% indicates that both techniques capture the productive macrostructure; however, the systematic disagreements reveal the greater sensitivity and specificity of the neural network.

The reclassification of 133 stands from the Medium class (guide-curve) to the Low class (ANN) is the result with the greatest practical impact of this study. This confirms the hypothesis that the guide-curve method, being univariate and focused only on dominant height, overestimates the potential of stands that, although they may have a reasonable height, present other limiting structural characteristics (e.g., high density, low diameter) (Skovsgaard & Vanclay, 2008). The ANN, by analyzing the complete dendrometric profile, identified these stands as more similar to the low-productivity class, providing a more realistic and accurate stratification of the forest potential.

Complementarily, the redefinition of the High Productivity Class (Class I) by the ANN, which proved to be more restrictive, demonstrates its ability to identify a profile of "productive efficiency" rather than just "superior height growth". Stands with high h_d but with a sub-optimal combination of other variables (such as basal area or diameter) were downgraded by the ANN to the Medium class. This refinement results in a more homogeneous and truly exceptional Class I, which is of fundamental importance for the management of high-value stands.

4.6 Limitations and future perspectives

Although the results of this study clearly demonstrate the advantages of the multivariate ANN approach, it is important to recognize that the classification was based exclusively on dendrometric variables obtained via forest inventory. The accuracy, spatial resolution, and robustness of site stratification could potentially be enhanced by the incorporation of additional explanatory variables that directly describe environmental factors or the current canopy condition.

Future research should focus on the synergistic integration of multiple data sources as inputs for the neural network. This includes edaphic and topographic attributes, which are determining factors of site quality at the local level. Additionally, the incorporation of high-resolution climate variables (Fiandino et al., 2020; Huang et al., 2024) and remote sensing data (Kanga, 2023) represents a promising avenue. Technologies such as multispectral satellite imagery (e.g., Landsat, Sentinel-2) can provide vegetation indices and texture metrics correlated to canopy vigor and structure (Gopalakrishnan et al., 2019), while LiDAR (Light Detection and Ranging) data offer detailed three-dimensional information on stand height, density, and structural complexity (Mensah et al., 2023; Gopalakrishnan et al., 2019).

Given the intrinsic capacity of SOM to handle high dimensionality, non-linearities, and data of different natures (Giraudel & Lek, 2001), it is expected that the fusion of these environmental and remote sensing variables with inventory data will result in even more precise, spatially explicit productive capacity classifications with greater generalization power, consolidating the ANN as a fundamental tool for precision silviculture and large-scale forest monitoring.

5. CONCLUSION

The SOM successfully mapped the underlying productivity gradient and the inverse relationship between stand density and individual tree growth. This resulted in a more accurate and silviculturally meaningful stratification than the univariate guide-curve method, notably due to increased sensitivity in detecting low-productivity stands—frequently misclassified as 'medium' by the traditional method—and higher specificity in identifying a more exclusive and high-performing high-productivity class. The statistical validity of these data-driven clusters was subsequently confirmed by discriminant analysis.

By providing a more reliable stratification, the ANN methodology offers a powerful strategic tool for improving forest management, enabling precision silviculture and enhancing yield forecasting. It is concluded that the application of Self-Organizing Maps represents a significant and validated methodological advance for site classification.

AUTHOR CONTRIBUTIONS

Silva, E. A.: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation, Writing - Review & Editing; Marques, P. A.: Software, Formal Analysis, Writing - Original Draft, Writing - Review & Editing; Azevedo, L. R. de: Writing - Review & Editing; Rabelo, G. S. P.: Writing - Review & Editing; Nogueira, G. S.: Writing - Review & Editing; Leite, H. G.: Conceptualization, Methodology, Writing – Review & Editing, Supervision; Oliveira, M. L. R. de: Conceptualization, Methodology, Writing – Review & Editing, Supervision, Project Administration.

DATA AVAILABILITY

The entire dataset supporting the findings of this study has been published within the article.

6. REFERENCES

Alvares, C. A., Stape, J. L., Sentelhas, P. C., de Moraes Gonçalves, J. L., & Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift, 22*(6), 711-728. https://doi.org/10.1127/0941-2948/2013/0507

Amaral, F. C. S. do; Santos, H. G. dos, Aglio, M. L. D., Duarte, M. N., Pereira, N. R., Oliveira, R. P. de, & Carvalho Junior, W. de (2004). *Mapeamento de solos e aptidão agrícola das terras do Estado de Minas Gerais.* Embrapa Solos.

Binoti, M. L. M. da S., Binoti, D. H. B., & Leite, H. G. (2013). Aplicação de redes neurais artificiais para estimação da altura de povoamentos equiâneos de eucalipto. *Revista Árvore, 37*(4), 639-645. https://doi.org/10.1590/S0100-67622013000400007

Campos, J. C. C., & Leite, H. G. (2025). *Mensuração Florestal: perguntas e respostas* (6ª ed). atualizada e ampliada. Editora UFV.

Clutter, J. L., Fortson, J. C., Pienaar, L. V., Brister, G. H., & Bailey, R. L. (1983). *Timber management: a quantitative approach.* John Wiley & Sons.

Cosenza, D. N., Leite, H. G., Marcatti, G. E., Binoti, D. H. B., Alcântara, A. E. M., & Rode, R. (2015). Classificação da capacidade produtiva de sítios florestais utilizando máquina de vetor de suporte e rede neural artificial. *Scientia Forestalis. 43*(108), 955-963.

Costa, J. A. F. (2010). Clustering and visualizing SOM results. In C. Fyfe, P. Tino, D. Charles, C. Garcia-Osorio, & H. Yin (Eds.), *Intelligent data engineering and automated learning – IDEAL 2010* (pp. 334–343). Springer. https://doi.org/10.1007/978-3-642-15381-5_41

Costa, J. A. F., Gonçalves, M. L., & Netto, M. L. A. (2011). Visualização e análise de agrupamentos usando redes auto-organizáveis, segmentação de imagens e índices de validação. *Journal of the Brazilian Neural Network Society, 9*(2), 91-103.

Duan, G., Lei, X., Zhang, X., & Liu, X. (2022). Site index modeling of larch using a mixed-effects model across regional site types in northern China. *Forests, 13*(5), 815. https://doi.org/10.3390/f13050815

Fiandino, S. I., Plevich, J. O., Tarico, J. C., Utello, M. J., Demaestri, M. A. & Gyenge, J. (2020). Modeling forest site productivity using climate data and topographic imagery in *Pinus elliottii* plantations of central Argentina. *Annals of Forest Science, 77*(4), 95. https://doi.org/10.1007/s13595-020-01006-3

García, O. (1995). Índices de sitio preliminares para eucalipto. *Ciencia & Investigación Forestal, 9*(1), 5–21. https://doi.org/10.52904/0718-4646.1995.211

Giraudel, J. L., & Lek, S. (2001). A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community analysis. *Ecological Modelling, 146*(1-3), 329-339. https://doi.org/10.1016/S0304-3800(01)00324-6

Gonçalves, A. C. (2022). Stand structure impacts on forest modelling. *Applied Sciences, 12*(14), 6963. https://doi.org/10.3390/app12146963

Gonçalves, J. L. M., Stape, J. L., Laclau, J. P., Smethurst, P., & Gava, J. L. (2004). Silvicultural effects on the productivity and wood quality of eucalypt plantations. *Forest Ecology and Management, 193*(1-2), 45-61. https://doi.org/10.1016/j.foreco.2004.01.022

Gonçalves, N., & Vigário, R. (2012). Clustering through SOM consistency. In A. Campilho & M. Kamel (Eds.), *Image analysis and recognition* (pp. 65–72). Springer. https://doi.org/10.1007/978-3-642-31295-3_8

Gopalakrishnan, R., Kauffman, J. S., Fagan, M. E., Coulston, J. W., Thomas, V. A., Wynne, R. H., Fox, T. R., & Quirino, V. F. (2019). Creating landscape-scale site index maps for the southeastern US is possible with airborne LiDAR and landsat imagery. *Forests, 10*(3), 234. https://doi.org/10.3390/F10030234

Gujarati, D. N., & Porter, D. C. (2011). *Econometria básica* (5ª ed.). Bookman.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Model assessment and selection. In *The elements of statistical learning.* Springer https://doi.org/10.1007/978-0-387-84858-7_7

Huang, L., Zhu, G., & Chen, G. (2024). Development of polymorphic index model for assessing subtropical secondary natural oak forest site quality under complex site and climate variables. *Forests, 15*(11), 1867. https://doi.org/10.3390/f15111867

Kanga, S. (2023). Advancements in remote sensing tools for forestry analysis. *Sustainable Forestry, 6*(1), 2269. https://doi.org/10.24294/sf.v6i1.2269

Kohonen, T. (2001). *Self-organizing maps*. Springer. http://dx.doi.org/10.1007/978-3-642-56927-2

Leisch, F., Jain, L. C., & Hornik, K. (1998). Cross-validation with active pattern selection for neural-network classifiers. *IEEE transactions on neural networks*, *9*(1), 35–41. https://doi.org/10.1109/72.655027

Leite, H., Castro, R., Silva, A., Júnior, C., Binoti, D., Castro, A. F., & Binoti, M. (2011). Classificação da capacidade orodutiva de oovoamentos de eucalipto utilizando diâmetro dominante. *Silva Lusitana, 19*(2), 181-195. https://scielo.pt/pdf/slu/v19n2/v19n2a04.pdf

Mensah, A. A., Jonzén, J., Nyström, K., Wallerman, J., & Nilsson, M. (2023). Mapping site index in coniferous forests using bi-temporal airborne laser scanning data and field data from the Swedish national forest inventory. *Forest Ecology and Management, 547*, 121395. https://doi.org/10.1016/j.foreco.2023.121395

Molina-Valero, J. A., Diéguez-Aranda, U., Álvarez-González, J. G., Castedo-Dorado, F. & Pérez-Cruzado, C. (2019). Assessing site form as an indicator of site quality in even-aged *Pinus radiata* D. Don stands in north-western Spain. *Annals of Forest Science, 76*(113). https://doi.org/10.1007/s13595-019-0904-1

Oliveira, M. L. R. de, Leite, H. G., Nogueira, G. S., Garcia, S. L. R., & Souza, A. L. de. (2008). Classificação da capacidade produtiva de povoamentos não desbastados de clones de eucalipto. *Pesquisa Agropecuaria Brasileira, 43*(11), 1559–1567. https://doi.org/10.1590/s0100-204x2008001100015

Penner, M., Woods, M., & Bilyk, A. (2023). Assessing site productivity via remote sensing—Age-independent site index estimation in even-aged forests. *Forests, 14*(8), 1541. https://doi.org/10.3390/f14081541

Prechelt, L. (2012). Early stopping—But when? In G. Montavon, G. B. Orr, & K.-R. Müller (Eds.), *Neural networks: Tricks of the trade* (Lecture Notes in Computer Science, Vol. 7700, pp. 53–67). Springer. https://doi.org/10.1007/978-3-642-35289-8_5

R Core Team. (2024). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Resende, R. T., Marcatti, G. E., Pinto, D. S., Takahashi, E. K., Cruz, C. D., de Resende, M. D. V., & de Resende, M. D. V. (2016). Intra-genotypic competition of *Eucalyptus* clones generated by environmental heterogeneity can optimize productivity in forest stands. *Forest Ecology and Management, 380*, 50–58. https://doi.org/10.1016/J.FORECO.2016.08.041

Rizzo-Martín, I., Hirigoyen-Domínguez, A., Arthus-Bacovich, R., Varo-Martínez, M. Á., & Navarro-Cerrillo, R. (2023). Site index estimation using airborne laser scanner data in *Eucalyptus dunnii* Maide stands in Uruguay. *Forests, 14*(5), 933. https://doi.org/10.3390/f14050933

Santos, P. P., Casas, G. G., Soares, C. P. B., Binoti, D. H. B., Lopes, L. S. de S., Marques, P. A., Magalhães, E. C., Araújo, A. C. da S. C., & Leite, H. G. (2025). Assessing sampling efficiency and artificial neural networks for tree height prediction and site index classification in forest stands. *Observatorio de la Economía Latinoamericana, 23*(3), e9409. https://doi.org/10.55905/oelv23n3-151

Scolforo, J. R. (1992). Curvas de índice de sítio para *Pinus caribaea* var. *hondurensis*. *IPEF,* (45), 40–47. https://www.ipef.br/publicacoes/scientia/nr45/cap06.pdf

Skovsgaard, J. P., & Vanclay, J. K. (2008). Forest site productivity: a review of the evolution of dendrometric concepts for even-aged stands. *Forestry, 81*(1), 13-31. https://doi.org/10.1093/forestry/cpm041

Skovsgaard, J. P., & Vanclay, J. K. (2013). Forest site productivity: A review of spatial and temporal variability in natural site conditions. *Forestry, 86*(3), 305–315. https://doi.org/10.1093/FORESTRY/CPT010

Sun, Y., Jin, X., Pukkala, T., & Li, F. (2022). A comparison of four methods for automatic delineation of tree stands from grids of LiDAR metrics. *Remote Sensing, 14*(24), 6192. https://doi.org/10.3390/rs14246192

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*. Springer. http:// doi.org/10.1007/978-0-387-21706-2

Vesanto, J., Himberg, J., Alhoniemi, E., & Parhankangas, J. (2000). SOM toolbox for MATLAB 5 documentation. In Proceedings of the MATLAB DSP Conference (pp. 35–40). Helsinki University of Technology.

Wehrens, R., & Kruisselbrink, J. (2018). Flexible self-organizing maps in kohonen 3.0. *Journal of Statistical Software, 87*(7), 1-18. https://doi.org/10.18637/jss.v087.i07

Weiskittel A. R, Hann D. W., Kershaw J. A. & Vanclay J. K. (2011) *Forest growth and yield modeling*. Wiley-Blackwell, Chichester

Yang, R., & Meng, J. (2022). Using advanced machine-learning algorithms to Estimate the Site Index of Masson Pine plantations. *Forests, 13*(12), 1976-1994. https://doi.org/10.3390/f13121976

Yu, L., Lai, K. K., Wang, S., & Huang, W. (2006). A bias-variance-complexity trade-off framework for complex system modeling. In M. Gavrilova, O. Gervasi, V. Kumar, H. Laganà, Y. Mun, D. Taniar, & C. J. K. Tan (Eds.), *Computational science and its applications – ICCSA 2006* (Lecture Notes in Computer Science, Vol. 3980, pp. 416–425). Springer. https://doi.org/10.1007/11751540_55