REVISTA FLORESTA
Engenharia Florestal

ISSN 1982-4688
Edição 54
Universidade Federal do Paraná
Setor de Ciências Agrárias

EXPLORING PATTERNS IN DENDROCHRONOLOGICAL DATA THROUGH CLUSTER ANALYSIS

Daniela Oliveira Silva Muraja[1*], Cecília Lemes Leite[1], Virginia Klausner[1], Alan Prestes[1], Iuri Rojahn da Silva

[1*]Universidade do Vale do Paraíba, Departamento de Física e Astronomia, São José dos Campos, São Paulo, Brasil - fys.dani@gmail.com ; leitelemescecilia@gmail.com ; viklausner@gmail.com ; aprestes@gmail.com ; iuri@univap.br

Received for publication: 18/06/2023 - Accepted for publication: 19/09/2023

Resumo

Explorando padrões em dados dendrocronológicos através da análise de cluster. Neste estudo, utilizou-se a metodologia do dendrograma para analisar séries temporais obtidas a partir da medição dos anéis de crescimento das árvores. Foram coletadas um total de 64 amostras de 21 árvores individuais. Polinômios foram aplicados para filtrar o padrão de crescimento natural das árvores e ampliar a influência de fatores externos, como as influências climáticas. A análise de agrupamento, usando a variância mínima de Ward e a distância euclidiana ao quadrado, foi utilizada para agrupar os dados com base na similaridade. Foram criados três dendrogramas, cada um contendo 10, 47 e 64 amostras, respectivamente. A análise revelou que as amostras com as maiores correlações, abrangendo mais de 95% do total de amostras, formaram grupos homogêneos. Além disso, a correlação de Pearson foi utilizada para confirmar os resultados obtidos nos dendrogramas. Portanto, pode-se afirmar que as amostras mais adequadas foram utilizadas na construção da cronologia média a partir dos dados disponíveis.

Palavras-chave: Dendrocronologia. Dendrograma. Anéis de Crescimento de Árvores.

Abstract

This study employed the dendrogram methodology to analyze time series data obtained from measuring tree growth rings. A total of 64 samples were collected from 21 individual trees. Polynomials were applied to filter the natural growth pattern of the trees and enhance the impact of external factors, such as climate influences. Cluster analysis using Ward's minimum variance and Euclidean squared distance was utilized to group the data based on similarity. Three dendrograms were constructed, consisting of 10, 47, and 64 samples, respectively. The analysis revealed that the samples with the highest correlations, encompassing over 95% of the total samples, formed homogeneous groups. Pearson correlation was also employed to confirm the results obtained from the dendrograms. Consequently, it can be affirmed that the most suitable samples were utilized in constructing the average chronology from the available data.

Keywords: Dendrochronology. Dendrogram. Tree Growth Ring.

INTRODUCTION

Cluster analysis is a powerful method widely used for studying large databases, with the main objective of grouping elements based on their shared characteristics. It utilizes a dissimilarity function to measure the distance between elements and form meaningful clusters. This analysis falls under the category of descriptive data mining, as it aims to explore the dataset and uncover properties that describe it (LINDEN, 2009).

The cluster analysis process involves iteratively grouping elements until all observations are merged into a single group (WILKS, 2006; KASSAMBARA, 2017). Determining the stopping point for the final analysis solution is a subjective decision that depends on the specific objectives of the analysis. Traditionally, researchers inspect the distances between clusters to identify the appropriate stopping point (WILKS, 2006; KASSAMBARA, 2017). Cluster analysis provides a hierarchical organization of data, and the resulting dendrogram or tree diagram visually represents the formation of grouped clusters. The dendrogram starts with each object in a separate group, and at each step, the two most similar clusters are combined into a new cluster, ensuring that they will never be separated.

In climatological studies, cluster analysis has various applications. It can be used to define climatic regimes based on airflow patterns, identify climatic regions using surface climate variables, and distinguish different atmospheric flow regimes (WILKS, 2006). For instance, Santos *et al.* (2021) applied cluster analysis to identify homogeneous precipitation regions in Paraná River to identify *El Niño*-Southern Oscillation effects.

In the context of dendrochronology, cluster analysis is employed to analyze time series data of tree growth rings, aiming to uncover spatial and temporal patterns of climate variations in response to natural forcings such as El Niño-Southern Oscillation (ENSO), volcanic eruptions, and solar cycles (ALLENDE, 2019; MURAJA, 2023). Tree growth rings serve as valuable records of the environmental and climate conditions that influenced their growth (BRIENEN, 2016). The variations in annual ring thickness reflect the sensitivity of trees to environmental factors, including temperature, light, and precipitation, which induce variations in cell size (ROIBU, 2020).

To ensure the reliability of the analysis, it is crucial to have tree samples that exhibit good similarity to each other. Dendrograms can aid in the identification of the most suitable samples for constructing average chronologies from the data (WILKS, 2006; KASSAMBARA, 2017). By utilizing cluster analysis, researchers can explore the hierarchical organization of data, gain insights into important concepts represented by the structure of clusters and unravel spatial and temporal patterns in climate. Previous studies in dendrochronology have extensively employed cluster analysis to create groups with homogeneous behavior (PIRAINO *et al.*, 2013; DOMÍNGUEZ-DELMÁS, 2014; MICHAEL, 2016; BABUSHKINA, 2016; LOVSETH, 2021).

In the present study, the agglomerative hierarchical process was utilized to perform the grouping analysis of dendrochronological series. The dissimilarity between elements was measured using the quadratic Euclidean distance, and the Ward's variance method was applied. The resulting groups were visualized using dendrograms or tree diagrams.

MATERIAL AND METHODS

Study samples

For this study, the forest species were chosen based on several factors. First, consideration was given to the morphological and anatomical characteristics of the wood, with a focus on the species' dendrochronological potential. This ensured that the chosen species had suitable characteristics for dendrochronological analysis. Secondly, the geographical position played a role in the selection process. The study focused on subtropical/temperate climate regions, which were deemed appropriate for the research objectives. Lastly, the age of the plants was considered. Trees of an appropriate age were chosen to ensure that the growth ring series captured a sufficient time span for analysis. Based on these criteria, the species *Ocotea porosa (Nees) Barroso* was chosen for the study. The samples were collected from an area located in the Southeast region of the state of Paraná, Brazil. Specifically, the *Ocotea porosa* samples were collected in the municipality of General Carneiro in January 2013. During the sample collection process, careful attention was given to checking for any signs of diseases or pest attacks on the trees. This thorough examination aimed to eliminate any potential influences that these factors could have on the growth ring series, ensuring the integrity and reliability of the collected data.

We collected a total of 64 samples from 21 different trees (individuals) using a non-destructive method (Pressler Borer/increment borer) at diameter to breast height (DBH - approximately 1.30 meters), this method was chosen since it does not cause any danger to the trees development. The laboratory preparation of the samples followed a systematic process:

(i) Samples were dried under shaded conditions.

(ii) The transverse surface of the samples was sanded and polished using 50-600 grit-sandpaper to enhance the visibility of tree rings.

(iii) Growth rings were carefully demarcated and were precisely measured using a VELMEX measuring table with a remarkable accuracy of 0.001 mm, ensuring that no false or missing rings were overlooked during the process.

The dating procedure was conducted from the tree bark to the pith. To ensure the accuracy of our tree-ring chronology, we implemented a cross-validation technique. This technique involved the use of a 50-year window to identify and rectify counting errors, as well as to detect false rings and partially missing rings.

Therefore, the dataset for this study consisted of time series data (mm x year) obtained by measuring the growth rings of 21 trees from a single species. A total of 64 samples were collected, with each growth ring representing one year of tree growth. It is important to note that this type of data is influenced by various limiting factors, as several parameters can affect plant growth and, consequently, the width of the tree's growth rings (COOK and KAIRIUKSTIS, 2013). When a tree is growing under optimal conditions, its rings tend to be wider, indicating favorable factors such as increased rainfall. Conversely, narrower rings indicate periods of stress or unfavorable conditions experienced by the plant.

To analyze the data accurately, it was necessary to address the natural growth tendency of the plant, which could interfere with the study. To mitigate this effect, trend removal was performed by subtracting polynomial fit functions using the Origin 7.0 software, as referenced in the user manual (Origin User's Manual). By applying this technique, the most suitable curve representing long-term growth rate patterns for each series was obtained. Subsequently, the calculated trends for each growth ring series were subtracted, resulting in standardized chronologies that reflected variations in ring width relative to their respective trends. This process transformed the growth ring series into stationary and comparable datasets. To achieve trend removal, linear, quadratic (degree 2 polynomial), and cubic (degree 3 polynomial) fit functions were employed. The selection of the most appropriate fit functions was based on their correlation with the time series data. This approach allowed for the preservation of the unique growth characteristics of each sample while retaining the climatological signals present in the data. For further details regarding the specific procedure described above, additional information can be found in Prestes

(2018) and Lorensi, Prestes (2016), Silva, *et al.* (2021) and Muraja, *et al.* (2023). Overall, the trend removal process ensured that the analysis focused on the variability in ring width rather than the overall growth trends of the trees, enabling a more accurate examination of the climatic and environmental influences on the growth patterns represented by the tree rings.

After obtaining the standardized chronologies by removing trends from the growth ring series, further analysis was conducted to identify the best-performing samples within each individual tree. Pearson correlation was performed between the samples extracted from the same tree. This correlation analysis helped in the selection of the most representative time series for each of the 21 trees. For each tree, the samples that exhibited a higher degree of similarity in their growth patterns were chosen to calculate the average chronology. This process aimed to ensure that the selected chronologies accurately represented the overall growth behavior of each individual tree. On the other hand, dendrochronological series with low adaptability, indicating limited similarity in their growth patterns, were discarded from further analysis. By performing the Pearson correlation and selecting the most suitable chronologies for each tree, the study aimed to enhance the quality and reliability of the dendrochronological data, ensuring that the chosen samples accurately reflected the growth characteristics of the respective trees. This selection process helped to improve the overall accuracy and representativeness of the dataset for subsequent analyses and interpretations.

Table 1: Information about the data that compose the time series of this study.
Tabela 1: Informação sobre os dados que compõem a série temporal deste estudo.

Tree	Samples	Years	1st Year
1	3	565	1446
2	2	84	1927
3	3	84	1927
4	2	151	1860
5	4	341	1670
6	3	83	1928
7	3	341	1670
8	3	91	1920
9	3	235	1776
10	4	379	1632
11	2	310	1701
12	5	225	1786
13	4	139	1872
14	2	194	1817
15	2	179	1832
16	4	491	1520
17	3	219	1792
18	4	278	1733
19	4	199	1812
20	1	239	1772
21	3	349	1662

Cluster analysis

The application of Cluster Analysis involves several essential steps to obtain meaningful results (GARCÍA; FONTI, 2008). These steps include:

1) Pre-processing the data: The input dataset must first be organized for the Cluster Analysis technique. This entails getting the data ready and making sure it can be used for analysis.

2) Decision similarity measure selection: The choice of an appropriate decision similarity measure is critical and is dependent on the study's specific goals. How the distance or dissimilarity between data points is calculated depends on the similarity measure.
3) Algorithm execution: The pre-processed data are next subjected to the Cluster Analysis algorithm. The agglomerative form of the algorithm was selected for this study. In the agglomerative technique, similar clusters are iteratively merged up until a stopping requirement is satisfied.
4) Selection of the best option: The type of data and the goals of the study will determine whether agglomerative or divisive clustering is the best option. In this instance, it was found that the agglomerative strategy was best suited to the study's data and goals.
5) Results evaluation and interpretation: After the Cluster Analysis algorithm has been used, the results must be assessed and understood. Analyzing the clusters that have formed, their traits, and any patterns or insights they have revealed is required for this. The interpretation of the data should be in line with the study's goals.

Since the thickness of the tree rings is measured in millimeters in the current study, the data scale utilized is regarded as quantitative. The data were standardized using the Squared Euclidean Distance to make analysis easier. By ensuring that all variables are on comparable scales, normalization helps prevent bias against variables with wider ranges. Following these stages enables the Cluster Analysis to explore and discover natural groupings or patterns within the dataset, offering insightful information for additional analysis and interpretation.

Hierarchical cluster analysis

Hierarchical Cluster Analysis is a powerful method that groups elements based on their similarity. It generates a hierarchy that captures different levels of grouping for each analyzed level. At the first level, all elements belong to a single cluster, and as we move towards the last level, the dataset is divided into a set of clusters. This hierarchical structure ensures that elements grouped together at higher levels remain in the same cluster at more specific levels. The analysis can be performed at different levels of the dendrogram based on the specific needs of the study.

There are two main types of algorithms used in Cluster Analysis, as described by Murtagh (1983):

1) Agglomerative: Each element starts as a single cluster, and pairs of clusters are merged based on their similarity until a single cluster is formed.
2) Divisive: Initially, all elements belong to a single cluster, which is then successively divided into smaller clusters until each cluster contains only one element or a stopping criterion is met.

To perform Cluster Analysis, a dissimilarity matrix is created, which captures the pairwise comparisons between elements. These comparisons are expressed as dissimilarity measures, typically calculated as distances between individuals. Statistical parameters are used to calculate the dissimilarity measures, which can be based on different criteria. The elements or groups with the smallest distances are merged into larger clusters, and the final result is often represented as a hierarchical dendrogram, facilitating the identification of groups (KASSAMBARA, 2017).

In the context described by Linden (2009), the dissimilarity or proximity between each pair of elements is captured in a matrix. Each entry in this matrix represents the degree of dissimilarity or proximity, denoted as "*pij*" Depending on the specific approach, the matrix can express either similarity or dissimilarity, but the fundamental concept remains the same. A transformation is then performed on the original data matrix:

$$r = \begin{bmatrix} x11 & \cdots & x1f & \cdots & x1p \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ xi1 & \cdots & xif & \cdots & xip \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ xn1 & \cdots & xnf & \cdots & xnp \end{bmatrix}$$

in a matriz of distances:

$$\mathrm{d} = \begin{bmatrix} 0 & & & & \\ d(2,1) & 0 & & & \\ d(3,1) & d(3,2) & 0 & & \\ \vdots & \vdots & \vdots & & \\ d(n,1) & d(n,2) & \cdots & \ldots & 0 \end{bmatrix} \begin{bmatrix} 0 & & & & \\ d(2,1) & 0 & & & \\ d(3,1) & d(3,2) & 0 & & \\ \vdots & \vdots & \vdots & & \\ d(n,1) & d(n,2) & \cdots & \ldots & 0 \end{bmatrix}$$

where *d(i,j)* is the distance between elements *i* and *j*.

Euclidian quadratic distance

The primary concept underlying cluster analysis is based on measuring the distance between data points. The objective is to form groups where the distances between the elements within each group are minimized, while the distances between the formed groups are maximized (WILKS, 2011).

In cluster analysis, the most commonly used distance measure for quantitative variables is the Euclidean distance in a K-dimensional space of the data vector (WILKS, 2006; KASSAMBARA, 2017). Specifically, the quadratic Euclidean distance is frequently employed, which calculates the sum of the squared differences between the values of two items. This distance, denoted as D.E, can be defined as follows:

$$D.E = \sum_{j=1}^{p} X_{ij} - Y_{ij}$$

Here, $X_{ij}X_{ij}$ and $Y_{ij}Y_{ij}$ represent the values of the j-th variable for items x and y, respectively. The quadratic Euclidean distance quantifies the dissimilarity between two objects. It is important to note that a smaller Euclidean distance indicates a higher similarity between objects being compared (BUENO; AGUIAR, 2004). As the distance approaches zero, the objects are considered more similar to each other.

By employing the Euclidean distance or its quadratic form, cluster analysis can effectively measure the dissimilarity between data points, enabling the formation of meaningful groups based on their similarity or proximity (VICINI, 2005).

Ward’s minimum variance method

In the field of dendrochronology, Ward's minimum variance method is a widely used hierarchical Cluster Analysis technique. Unlike other methods, Ward's method does not directly operate on the distance matrix. Instead, it starts with individual data points assigned to separate groups and gradually merges pairs of groups until all elements are united in a single group, typically after n-1 steps (BUENO; AGUIAR, 2004; VICINI; SOUZA, 2005; WILKS, 2011).

The key criterion for merging groups in each step is to minimize the sum of squared distances between the data points and the centroids of their respective groups. This process involves recalculating the group mean (or centroid) by incorporating data from the previously separated groups and then computing the squared distances accordingly (WILKS, 2006; KASSAMBARA, 2017).

In essence, Ward's minimum variance method aims to minimize the variations within groups by effectively combining similar data points. By iteratively merging groups based on this criterion, the method provides insights into the hierarchical structure and relationships among the data points in dendrochronological studies (WILKS, 2006; KASSAMBARA, 2017).

Dendrogram

Using a dendrogram, or tree diagram, to represent the process and outcomes of cluster analysis is a popular practice. A 5-point cluster is shown in a simplified dendrogram in Figure 1. On the left side of Figure 1, where each point is originally treated as a separate cluster, the analysis starts. The level of similarity is measured along the horizontal axis using the Euclidean distance metric, and different observations are listed along the vertical axis.

Points X3 and X4 are combined to create a new cluster in the first stage. The vertical bar that connects the two locations visually depicts the separation between them, with the length of the bar corresponding to the distance value of d3,4 (WILKS, 2006; KASSAMBARA, 2017). Points X1 and X2 are combined into another cluster at the following stage since their distance is the shortest among the six calculated from the four previously identified clusters. Figure 1's right side displays the distance value d1,2 which, according to Wilks (2006), is less than d3,4.

In the final step, point X5 is joined with the cluster formed by the pair (X3, X4), resulting in a two-stage hierarchy. The dendrogram visually represents this merging process, offering insights into the relationships and distances between the data points (WILKS, 2006; KASSAMBARA, 2017).

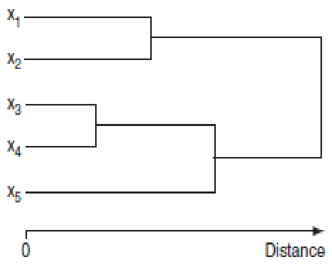

Figure 1: Illustration of a dendrogram. The results of the four clustering steps are indicated as the original five lines are progressively joined from left to right, with the distances joined clusters indicated by the positions of the vertical lines.

Figura 1: Ilustração de um dendrograma. Os resultados dos quatro passos da análise de Cluster são indicados à medida que as cinco linhas originais se juntam progressivamente da esquerda para a direita, com as distâncias unidas cluster é indicado pela posição das linhas verticais.
Source: Wilks (2006)
Fonte: Wilks (2006).

The dendrograms for this study were made using the Linkage and Dendrogram tools from MATLAB Software. Tools to generate dendrogram visualizations and calculating linkage matrices are included in these packages. The Linkage package (https://www.mathworks.com/help/stats/linkage.html), which depicts the hierarchical relationships between data points based on distance measures, offers methods for computing linkage matrices. Dendrograms are constructed using these matrices as their basis. The Dendrogram package (https://www.mathworks.com/help/stats/dendrogram.html) makes it straightforward to generate dendrograms when utilizing the computed linkage matrices.

The dendrograms in this study were effectively constructed by utilizing the capabilities of these MATLAB packages, enabling a clear depiction of the hierarchical clustering structure, and assisting in the interpretation of the data.

RESULTS

It is well-known that the standard procedure for obtaining a dendrochronological series for a given region involves the following steps:

1. Sample Collection: Initially, samples are collected from trees within the region of interest. This typically involves using an auger or increment drill to extract a cylindrical wood sample from the tree trunk. Collecting samples from multiple trees is crucial to obtain a representative sample of the variation in growth conditions.

2. Sample Preparation: The collected samples undergo meticulous preparation for analysis. This includes carefully drying the samples under controlled conditions to eliminate moisture and ensure they are ready for examination.

3. Series Dating: Samples are then meticulously sanded or polished to enhance the visibility of growth rings. Each growth ring corresponds to one year of the tree's growth. The counting of rings begins from the center of the sample and proceeds outward to determine the age of each ring.

4. Ring Width Measurement: To construct a dendrochronological series, the widths of the rings in each sample are measured with a high degree of precision. Specialized measuring devices are often employed for this purpose.

5. Standardization: To facilitate the comparison of samples from different trees and create an averaged time series, the ring width data undergoes a standardization process. This entails transforming the data to remove individual variations in tree growth rates.

6. Statistical Analysis: Following standardization, statistical analyses are performed to identify common growth patterns among the trees. These analyses may involve the utilization of techniques such as regression, or other statistical methods.

7. Development of Dendrochronological Series: Based on the outcomes of the statistical analysis, an averaged dendrochronological series is generated for the region. This series represents annual fluctuations in tree growth over time.

8. Interpretation and Use: The final dendrochronological series finds application in various fields, including climate studies, paleo-reconstructions, and analyses of environmental trends.

It is vital to emphasize that the quality of a dendrochronological series depends on the careful choice of trees, sampling techniques and the accuracy of ring measurements. Additionally, interpreting dendrochronological data necessitates a solid understanding of tree ecology and the growing conditions specific to the region under investigation.

The importance of this study lies in presenting an efficient alternative for deriving an average chronology, which can serve as another possibility to the procedures outlined in steps 6 and 7 above. To this end, cluster analysis will be conducted using Ward's minimum variance and squared Euclidean distance to group data based on similarity. With the results in hand, we will be able to determine which samples are most suitable for constructing the average chronology from the available data.

Cluster analysis was used to judge the accuracy of the chronology that was acquired. The separate series were split into three intervals of equal length after being normalized by eliminating trends. This division was necessary due to the different lengths of the series. The groups that formed were as follows: the first interval, which had 10 samples, spanned the years 1670 to 1832; the second, the years 1832 to 1935; and the third, the years 1935 to 2011 with all 64 samples.

Three dendrograms were created to show the clustering tendencies within each period. Figures 2, 3, and 4 show these dendrograms, which illustrate the hierarchical links between the various series within each time period. In these dendrograms, the vertical axis represents the tree's radius, denoted by the letter 'R,' with numbers indicating the specific tree as per Table 1, arranged from the oldest to the youngest. The letters 'A,' 'B,' 'C,' and so on along the vertical axis signify the number of rings in the sample. Meanwhile, the horizontal axis reflects the Euclidean distance metric. The dendrograms help in evaluating and interpreting the chronology quality by revealing parallels and differences in the growth patterns of the trees over these particular time periods.

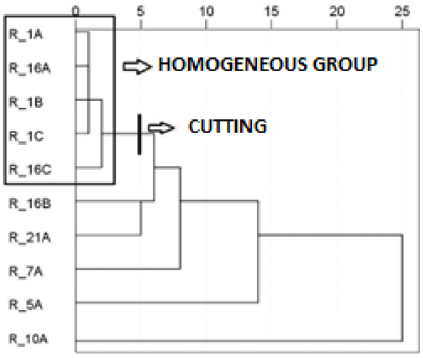

Figure 2: Dendrogram of the series of growth rings in the period 1670 to 1832 constructed with 10 rays obtained from 6 trees.

Figura 2: Dendrograma das séries de anéis de crescimento do período de 1670 até 1832 construídas amostras obtidas de 6 árvores.

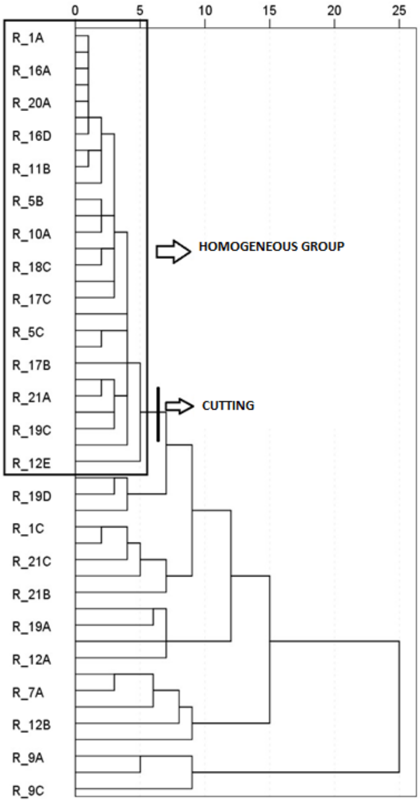

Figure 3: Dendrogram of the growth ring series in the period 1832 to 1935, constructed with 47 samples obtained from 16 trees.

Figura 3: Dendrograma da série de anéis de crescimento no período de 1832 até 1935, construído com 47 amostras obtidas de 16 árvores.

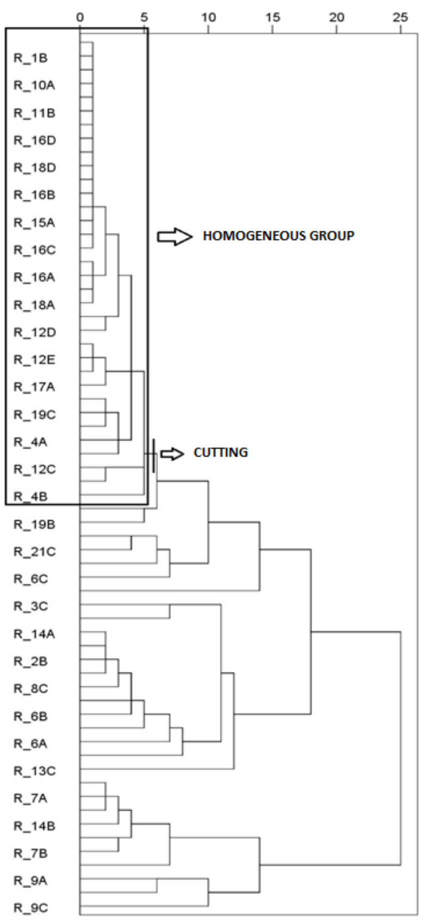

Figure 4: Dendrogram of the growth ring series from 1935 to 2011, built with 64 samples obtained from 21 trees.
Figura 4: Dendrograma da série de anéis de crescimento de 1935 até 2011, construída com 64 amostras obtidas de 21 árvores.

The chronology for the specific location, spanning a period of 565 years, was obtained from the dendrograms. The resulting chronology, shown in Figure 5, offers an extensive overview of the growth trends noticed over time. The number of samples included in each period is shown on the red line in Figure 5, which symbolizes the chronology and reflects the data used for building the series.

Researchers can learn a lot about the growth dynamics and environmental influences on the researched trees by examining the chronology. The visual representation of temporal differences in tree development in the given picture emphasizes the significance of the chosen samples from various time periods in capturing the broad trends and patterns shown in the dendrochronological series.

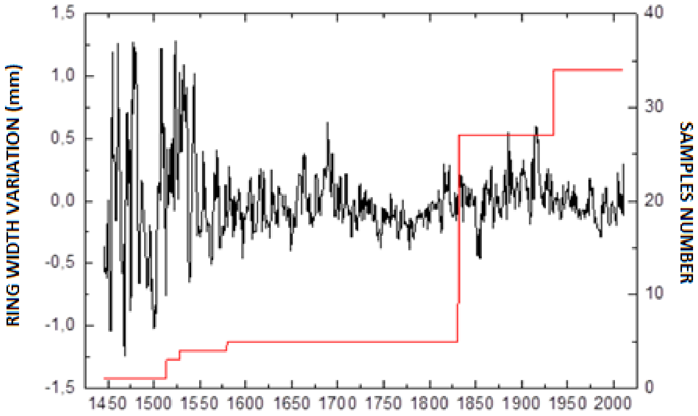

Figure 5: Mean chronology (black line) and number of samples (red line) of trees used in its construction.
Figura 5: Cronologia média (linha preta) e número de amostras (linha vermelha) de árvore usados nesta construção.

DISCUSSÃO

The Fenon line, a cut line with a rescaled Euclidean distance of 5, was used to pick the groupings. The individual needs of the researcher are used to determine the position of the Fenon line. Its value can vary up to a maximum distance of 12.5, which accords with the Euclidean average (LOPES, 2004).

The creation of the average chronology was carried out after the identification of homogenous groupings in each dendrogram. The resulting time series spans a period of 1446 years, beginning with the date of the oldest tree and ending with 2011, the year that the final ring was formed. The series for the time period prior to 1670, which included just two samples, was created by averaging those samples.

A homogenous group was created in Figure 2 by combining the data from trees 1 and 16, which make up 50% of the samples utilized in the dendrogram and reflect the interval including the oldest trees.

It was found that a homogenous group of 27 samples, or 57% of the samples contained in the dendrogram, can be found in Figure 3, which displays the dendrogram for the second interval. The average series was made up of the trees 10, 11, 13, 16, 18, and 20, whereas 1, 5, 7, 12, 17, 19, and 21 made up only a portion of the series.

The homogeneous group for creating the average time series in Figure 4's dendrogram of the third interval was made up of 34 of the total 64 samples. Seven trees provided all of their samples, seven trees participated in part of the study, and seven trees were wholly cut off from the group. There are no trees in this group that are younger than 100 years old.

A comparison between the three dendrograms reveals a consistent behavior in the time series. Out of the five samples forming the homogeneous group in the first dendrogram, four samples remained in the other two dendrograms. Similarly, of the 27 samples grouped in the second dendrogram, 24 samples persisted in the third dendrogram.

CONCLUSÕES

From the results obtained, this study draws the following conclusions:

- Identification of Optimal Growth Rings: Cluster analysis proved to be an effective method for selecting the best growth rings among the tree samples. By grouping similar samples together, it facilitated the identification of key growth patterns and characteristics.
- Construction of Chronology: The dendrograms generated from the cluster analysis provided a chronology that can be utilized for various future analyses, such as dendroclimatology and paleoclimatology studies. The chronology offers valuable insights into past climate patterns and can contribute to a better understanding of historical environmental conditions.
- Comparative Analysis of Time Series: Cluster analysis demonstrated its utility in comparing time series with different sizes. By clustering and grouping the data, it allowed for the examination of similarities and differences among the series, enabling researchers to identify patterns and trends.
- Selecting Subsamples and Evaluating Homogeneity: The dendrograms derived from cluster analysis facilitated the selection of subsamples from the series. It provided a means to discard samples that did not exhibit strong correlations and allowed for the assessment of homogeneity among different subsets. This capability enhances the precision and accuracy of subsequent analyses.
- Methodological Consistency: The analysis confirmed that the applied cluster analysis method is a robust and reliable approach, as more than 95% of the samples exhibiting the highest correlations formed homogeneous groups. This indicates the consistency and validity of the results obtained through the methodology.

Overall, this study shows how useful cluster analysis may be as a tool for dendrochronological studies. It allows for the determination of the best growth rings, the creation of chronologies, the comparison of time series, the choice of subsamples, and the assessment of homogeneity. These discoveries enrich the study of dendroclimatology and paleoclimatology by giving important new information about historical environmental conditions and climate dynamics.

REFERÊNCIAS

ALLENDE, T. C.; MACIAS, J. L.; MENDOZA, M. E.; DIAZ, J. V. Evidence of volcanic activity in the growth rings of trees at the Tacaná volcano, Mexico–Guatemala border. *Canadian Journal of Forest Research*, v. 50, p. 65-72, 2019. https://doi.org/10.1139/cjfr-2019-0214

BABUSHKINA, E. A.; VAGANOV, E. A.; GRACHEV, A. M.; ORESHKOVA, N. V.; BELOKOPYTOVA, L. V.; KOSTYAKOVA, T. V.; KRUTOVSKY, K. V. The effect of individual genetic heterozygosity on general homeostasis, heterosis and resilience in Siberian larch (Larix sibirica Ledeb.) using dendrochronology and microsatellite loci genotyping. Dendrochronologia, v. 38, p. 26-37, 2016.

BUENO, Y. M.; AGUIAR, D. R. Determinantes do grau de envolvimento na exportadora e suas implicações nas exportações brasileiras de carne de frango. Gestão e Produção, v. 11, n. 2, p. 211–220, 2004.

BRIENEN, R. J. W.; SCHÖNGART, J.; ZUIDEMA, P. A. Tree Rings in the Tropics: Insights into the Ecology and Climate Sensitivity of Tropical Trees. In: Goldstein, G., Santiago, L. (eds) Tropical Tree Physiology. Tree Physiology, vol 6, 2016. https://doi.org/10.1007/978-3-319-27422-5_20

COOK, E. R.; KAIRIUKSTIS, L. A. Methods of Dendrochronology: Applications in the Environmental Sciences. Ed. 1. eBook ISBN: 978-94-015-7879-0. 2013. https://doi.org/10.1007/978-94-015-7879-0

DOMÍNGUEZ-DELMÁS, M.; DRIESSEN, M.; GARCÍA-GONZÁLEZ, I.; HELMOND N. V.; VISSER, R.; JANSMA, E. Long-distance oak supply in mid-2nd century AD revealed: the case of a Roman harbour (Voorburg-Arentsburg) in the Netherlands, Journal of Archaeological Science, v. 41, p. 642-654, 2014. https://doi.org/10.1016/j.jas.2013.09.009.

GARCÍA, G. I.; FONTI, P. Ensuring a representative sample of earlywood vessels for dendroecological studies: an example from two ring-porous species. Trees, v. 22, n. 2, p. 237-244, 2008.

HASELEIN, C. R.; SANTINI. E. J.; LONGHI, S. J.; ROSSO, S.; FERNANDES, D. L. G.; MENEZES, L. F. Agrupamento de árvores matrizes de *eucalyptus grandis* em função das variáveis dendrométricas e das características tecnológicas da madeira. Ciência Florestal, v. 14, n. 2, p. 133–144, 2014.

KASSAMBARA, A. Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning. (Multivariate Analysis Book 1). Editora sthda.com, 2017.

LINDEN, R. Técnicas de agrupamento, v. 1, 2009.

LORENSI, C.; PRESTES, A. Dendroclimatological Reconstruction of Spring-Summer Precipitation for Fazenda Rio Grande, PR, with samples of *Araucaria angustifolia (Bertol.) Kuntze*. Revista Árvore, scielo, v. 40, p. 347 – 354, 2016.

LOVSETH, J. T.; GRONINGER, J. W.; RUFFNER, C. M.; BOOKOUT, B.; KELLER, T.; SILCOX, K.; SCHNAITMAN, D. Using Floristic Analysis and Dendrochronology to Plan Ecological Restoration within a Forest–Prairie Ecotone, Natural Areas Journal, v. 41, p. 238-248, 2021.

MICHAEL, S. Finite Mixture Modeling of Gaussian Regression Time Series with Application to Dendrochronology. *J Classif*, v. 33, p. 412–441. 2016. https://doi.org/10.1007/s00357-016-9216-4

MURAJA, D. O. S.; KLAUSNER, V.; PRESTES, A.; DA SILVA, I. R. Ocean-atmosphere interaction identified in tree-ring time series from Southern Brazil using cross-wavelet analysis. Theoretical & Applied Climatology, 2023.

MURTAGH, F. A survey of recent advances in hierarchical clustering algorithms. The Computer Journal, p. 354–359, 1983.

NCSS. Hierarchical Clustering/Dendrograms. 2019. Available at: https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Hierarchical_Clustering-Dendrograms.pdf

PIRAINO, S.; CAMIZ, S.; FILIPPO, A.; PIOVESAN, G.; SPADA, F. A dendrochronological analysis of *Pinus pinea L.* on the Italian mid-Tyrrhenian coast. Geochronometria, v. 40, n. 1, p. 77-89, 2013.

PRESTES, A.; KLAUSNER, V.; DA SILVA, I. R.; OJEDA-GONZÁLEZ, A.; LORENSI, C. Araucaria growth response to solar and climate variability in South Brazil. Annales Geophysicae, v. 36, p. 717-729, 2018.

ROIBU C-C.; SFECIA, V.; MURSA, A.; IONITA, M.; NAGAVCIUC, V.; CHIRILOAEI, F.; LESAN, I.; POPA, I. The Climatic Response of Tree Ring Width Components of Ash (*Fraxinus excelsior* L.) and Common Oak (*Quercus robur* L.) from Eastern Europe. *Forest*, 2020. https://doi.org/10.3390/f11050600

SANTOS, E.; DE FREITAS, E. D.; RAFEE, S. A. A.; FUJITA, T.; RUDKE, A. P.; MARTINS, L. D.; DE SOUZA, R. A. F.; MARTINS, J. A. Spatio-Temporal Variability of Wet and Drought Events in the Paraná River Basin—Brazil and Its Association with the *El Niño*—Southern Oscillation Phenomenon. *International Journal of Climatology*, v. 41, n. 10, 2021, p. 4879-4897, https://doi.org/10.1002/joc.7104. Accessed 26 Jul. 2023.

SILVA, D. O.; KLAUSNER, V.; PRESTES, A.; MACEDO, H. G.; AAKALA, T.; DA SILVA, I. R. Principal components analysis: An alternative way for removing natural growth trends. Pure and Applied Geophysics, 2021.

VICINI, L.; SOUZA, A. M. Análise multivariada da teoria à prática. Santa Maria: UFSM, CCNE, 2005.

WILCZYŃSKI, S.; FELIKSIK, E. Local chronologies and regional diversity of dendrochronological signal of Douglas fir in Poland. Geochronometria, v. 26, n. 1, p. 69-80, 2007.

WILKS, D. S. Cluster analysis. In: Statistical Methods in the Atmospheric Sciences. 2nd edition. ed. [S.l.]: Elsevier, 2016.

WILKS, D. S. Statistical methods in the atmospheric sciences. Academic press, 2011.