ISSN: 2447-3359

REVISTA DE GEOCIÊNCIAS DO NORDESTE

Northeast Geosciences Journal

v. 8, nº 1 (2022)

https://doi.org/10.21680/2447-3359.2021v8n1ID24096

Principal component analysis applied to dendrochronology

Análise de componentes principais aplicada à dendrocronologia

Daniela Oliveira da Silva[1]; Virginia Klausner[2]; Alan Prestes[3]; Humberto Gimenes Macedo[4]

[1] Vale do Paraíba University, São José dos Campos/SP, Brazil. Email: fys.dani@gmail.com
ORCID: https://orcid.org/0000-0002-3851-3368
[2] Vale do Paraíba University, São José dos Campos/SP, Brazil. Email: viklausner@gmail.com
ORCID: https://orcid.org/0000-0003-0250-5574
[3] Vale do Paraíba University, São José dos Campos/SP, Brazil. Email: aprestes@gmail.com
ORCID: https://orcid.org/0000-0001-9403-8103
[4] Vale do Paraíba University, São José dos Campos/SP, Brazil. Email: gimeneshumberto@outlook.com
ORCID: https://orcid.org/0000-0002-0858-6283

Abstract: This work uses samples of the species *Ocotea porosa (Nees & Mart) Barroso* (Imbuia), collected in the city of General Carneiro, Southeast region of the State of Paraná (26º24'01 25"S 51º24'03 91"W), Brazil, to generate average chronology (GC index) of this region. The objective of this article is to remove the natural growth trends of trees using a tool that is still little explored for this purpose, Principal Component Analysis (PCA). In each tree sample, the width of each growth ring was measured, obtaining a time series (1 ring per year). The samples were selected using Cluster Analysis, which classifies samples based on their similarities. Once the Principal Components (PCs) were obtained, the dendrochronological series were reconstructed without the first PC. This methodology is an estimate of the trend that best represents the natural growth of all trees on the site. The arithmetic mean of the series without the 1st PC is the GC index. It was found that PCA has three benefits: fast data processing, preservation of low-frequency signals and, when integrated with a powerful tool, the Alternated Least Squares (ALS) method, missing data estimation.

Keywords: Dendrochronology; Principal Component; Natural Records.

Resumo: Este trabalho utiliza amostras da espécie *Ocotea porosa (Nees & Mart) Barroso* (Imbuia), coletadas na cidade de General Carneiro, região Sudeste do Estado do Paraná (26º24'01 25"S 51º24'03 91"O), Brasil, para gerar cronologia média (index GC) desta região. O objetivo deste artigo é remover as tendências de crescimento natural das árvores utilizando uma ferramenta ainda pouco explorada para este propósito, a Análise de Componentes Principais (PCA). Em cada amostra de árvores, foi medida a largura de cada anel de crescimento, obtendo-se uma série temporal (1 anel por ano). As amostras foram selecionadas por meio de Análise de Clusters, que classifica amostras com base em suas semelhanças. Obtidas as Componentes Principais (PCs), reconstruiu-se as séries dendrocronológicas sem a primeira PC. Essa metodologia é uma estimativa da tendência que melhor representa o crescimento natural de todas as árvores do local. A média aritmética das séries sem a 1ª PC é o index GC. Verificou-se que a PCA traz três benefícios: processamento rápido de dados, preservação dos sinais de baixa frequência e, quando integrada a uma ferramenta poderosa, o método dos mínimos quadrados alternados (ALS), estimativa de dados faltantes.

Palavras-chave: Dendrocronologia; Componente Principal; Registros Naturais.

Recebido: 22/02/2021; Aceito: 11/04/2022; Publicado: 14/06/2022.

1. Introduction

Dendrochronology is the name given to the study of tree growth rings. This type of technique allows us to infer what the environmental conditions were like at the time of formation of each ring. The pioneering in this area is given to the astronomer Andrew Ellicott Douglass who noticed around 1900, variations in the widths of the rings in some trees. These variations were attributed to the influence of external factors (climatic conditions, solar activity, nutrients in the soil, among others) and internal factors (natural growth tendency of a certain species). Therefore, trees are important records of a variety of natural processes in the environment and changes caused by humanity, such as pollution and contamination (SPEER, 1971).

Dendrochronological data are used to analyze the spatial and temporal patterns of climate variation in the face of natural forcing (FYE; CLEAVELAND, 2001). This method consists of determining the age of the trees by analyzing their rings formed in their years of growth, as well as extracting other information, such as the occurrence of climatic events (SCHWEINGRUBER, 1988). This technique is also used in the study of the sunspot cycle. Several studies observed that the widths of the growth rings of different species were related to the periods of minimum and maximum solar activity (DERGACHEV; RASPOPOV, 2000; RASPOPOV et al., 2001; RIGOZO et al., 2004).

There are several methods for data analysis and the choice of method depends on the data characteristics and what is desired to be highlighted by the analysis in question. In dendrochronology, it is common to have a great quantity of data, that is, many samples of trees and, consequently, measured rings, with many annual rings. Therefore, it is necessary to choose a statistical analysis that helps to visualize this large volume of data simply, but with the least possible loss of information.

Principal Component Analysis (PCA) demonstrates a great ability to work with big data. According to Anderson (2003), PCA consists of linear combinations of variables with special properties concerning their variances. In this article, time series with different sample sizes will be used, therefore, it is necessary to work with the Alternating Least Squares (ALS) method. The ALS technique performs an extrapolation of the data based on the originals. That is, the oldest samples are used to estimate the missing values and, therefore, all samples will have the same size. The principal components (PCs), calculated by the PCA, facilitate the discrimination of the internal and external factors that acted in the growth of the group of trees. Thus, it is possible to evaluate geophysical and climatic factors that influence the climate of a given region and that were recorded in the development of each growth ring. In other words, the growth response of each species can be associated with environmental factors such as temperature, precipitation, relative humidity, atmosphere-ocean relationship etc., and consequently through these responses, climatic reconstructions can be carried out. These reconstructions will depend on the age of the chronology of the study region.

The present study aims to show the advantage of applying the PCA methodology in the generation of dendrochronological series in an automated way. The traditional technique of dendrochronology performs the removal of signals of influences internal to the growth of the tree of each one of the samples through iterative means, being more computationally expensive. The removal of the natural growth trend as a first step is fundamental for the analysis of external factors, such as the study of the climate of the region where the tree samples were collected. In addition, this article presents the mathematical step-by-step PCA application of an example case which helps in understanding the method.

2. Methodology

2.1 Obtaining dendrochronological data

Dendrochronological data were obtained by measuring the thickness of each growth ring. Each ring represents the growth variation during the one year of the evaluated tree. After carrying out all the measurements, a time series, i.e., the dendrochronological series, is obtained. The average dendrochronology (Index) for the study region is obtained through the arithmetic mean of the widths of the rings of the different dendrochronological series (COOK; KAIRIUKSTIS, 1990). For the calculation of the index, it is advisable to consider as many trees as possible within a given region.

The growth rings of some tree species have light and dark areas, as seen in Figure 1. The light bands represent the wood formed at the beginning of the growth period (earlywood), in which they have thin cellulosic walls and large cytoplasmic diameters. The dark parts are related to wood produced at the end of the growth period (latewood) and have thicker cell walls and smaller cytoplasmic diameters (LISI, 2000).

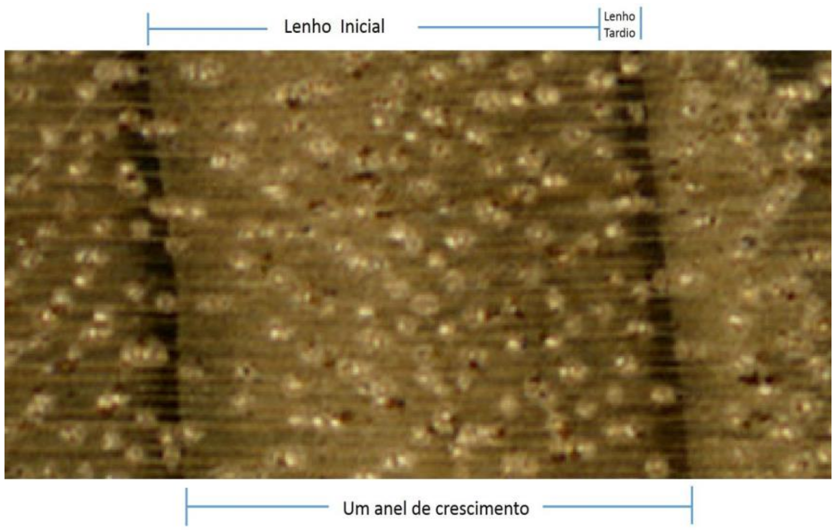


Figure 1 - Structure of a growth ring of the species Ocotea porosa (Nees & Mart) Barroso in which it is possible to observe early and latewood.
Source: Authors (2020).

The dataset used in this work is a series of growth rings of trees of the species *Ocotea porosa (Nees & Mart) Barroso*, known as Imbuia, from the Southeast region of the state of Paraná, municipality of General Carneiro (Figures 2a and 2b). Sixty-four samples were obtained from 21 trees, collected in January 2013 by a non-destructive method, with the aid of Pressler probes, at the height of the DBH (1.30 m). The number of samples obtained from each tree is shown in Table 1.

a) 

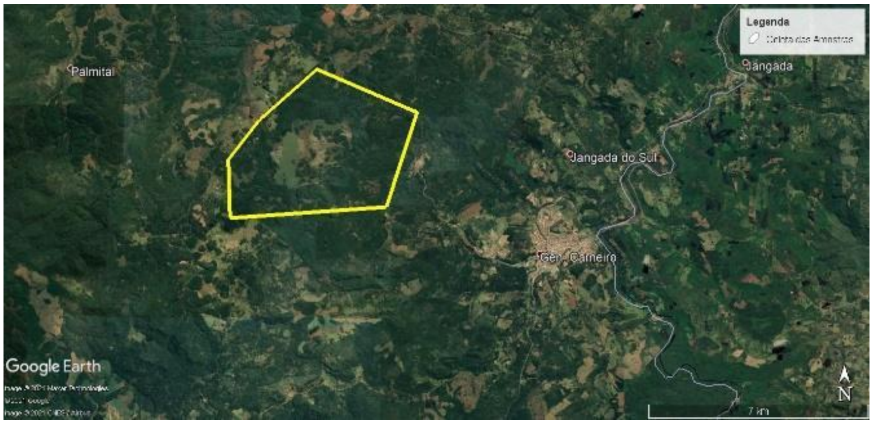
b) 

Figure 2 – Maps of the collection region. a) General map of the southern region of Brazil with the marking of the city of General Carneiro. b) Sample collection site in the city of General Carneiro.
Source: Google Earth Pro (2020).

Table 1 – Number of samples obtained per tree.

Árvore	1	2	3	4	5	6
Amostras	3	2	3	2	4	3
Árvore	7	8	9	10	11	12
Amostras	3	3	3	4	2	5
Árvore	13	14	15	16	17	18
Amostras	4	2	2	4	3	4
Árvore	19	20	21			
Amostras	4	1	3			

Source: Authors (2020).

The samples were initially polished on the transverse surface with different sandpaper (from 50 to 600 grit). To measure the growth rings, a VELMEX measuring table was used, with a precision of 0.001 mm, connected to a stereoscopic microscope with a micrometric reticle of the Laboratory of Natural Records, Universidade do Vale do Paraíba - UNIVAP.

The dating was done from the bark to the pith. The last ring was formed until the collection date corresponding to the year 2011. In the southern region of Brazil, the beginning of the formation of a growth ring occurs in Spring/Summer and, therefore, the ring corresponding to the year 2012 still had not completed its growth in January 2013 (period of collection).

The choice of samples was made using the agglomerative hierarchical process, with the measure of dissimilarity between the elements (Cluster Analysis) given by the Quadratic Euclidean distance, using Ward's variance method (RICKEN et al., 2018). This procedure is important so that only trees with similar growth enter the study, thus avoiding increasing the built-in error. So, of the 64 samples, only 37 were chosen, which are listed in Table 2.

Table 2 – Trees and their respective samples selected by Cluster Analysis.

Tree	Sample (s)	Tree	Sample (s)
1	A, B e C	15	A
4	A e B	16	A, B, C e D
5	B e C	17	A, B e C
7	C	18	A, B, C e D
10	A, B, C e D	19	C e D
11	A e B	20	A
12	A, B, C, D, e E	21	A
13	A e D		

Source: Authors (2020).

2.2 Mathematical Application of PCA

In this section, the mathematical application of PCA will be treated with a simple numerical example.

The PCs are obtained by diagonalizing semi-defined positive symmetric matrices. There are several programs capable of matrix calculations such as diagonalization (HONGYU; SANDANIELO; OLIVEIRA JUNIOR, 2016). The numerical example provided by Gonzalez and Woods (2000), and presented below, illustrates obtaining the PCs.

Covariance (or variance) represents the degree to which each variable is linearly correlated with the other. The equation for calculating two-dimensional data is:

$$cov(x,y) = \frac{\Sigma_{i-I}^{n}\{(xi-\underline{x})\cdot(yi-\underline{y})\}}{n} \quad \textbf{(Equation 1)}$$

x and y being data lists, with x being the first dimension and y being the second dimension; and are the averages of the lists; xi and yi are each of the elements of the lists, in the i-th position; n is the number of data items. When the data start at index 0 (representing a sample), in the denominator and sum we use n-1, if it is the total set, we use n itself. For data with more than two dimensions, covariance (Equation 1) is performed for each pair of dimensions. If it is a matrix of three dimensions (x, y, and z), we have the following covariance matrix:

$$matrizcov = \begin{pmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{pmatrix} \qquad \textbf{(Equation 2)}$$

For dendrochronological data, each dimension of the matrix is given by a sample from a different tree. From the main diagonal of this matrix, the variances are obtained, and in the other positions, the correlations between the directions are obtained. As this matrix is symmetric and real, one can always find a set of orthonormal eigenvectors (ANTON; RORRES, 2004) that are the main axes of the matrix.

Considering M samples of vectors, the average vector (mx) can be calculated by:

$$mx = \frac{1}{M}\Sigma_{i=1}^{M} xi \qquad \textbf{(Equation 3)}$$

As an example, we will use a set of 3D vectors with 4 samples for each step of these calculations. Assuming the following values for each of the 4 samples:

$$x1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad \textbf{(Equation 4)}$$

$$x2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad \textbf{(Equation 5)}$$

$$x3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \qquad \textbf{(Equation 6)}$$

$$x1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \qquad \textbf{(Equation 7)}$$

We will have the following average vector:

$$mx = \begin{bmatrix} 3/4 \\ 1/4 \\ 1/4 \end{bmatrix} \qquad \textbf{(Equation 8)}$$

Each xi is subtracted from the mean vector (mx) to calculate the covariance matrix:

$$x1 - mx = \begin{bmatrix} -3/4 \\ -1/4 \\ -1/4 \end{bmatrix} \qquad \textbf{(Equation 9)}$$

$$x2 - mx = \begin{bmatrix} 1/4 \\ -1/4 \\ -1/4 \end{bmatrix} \qquad \textbf{(Equation 10)}$$

$$x3 - mx = \begin{bmatrix} 1/4 \\ 3/4 \\ -1/4 \end{bmatrix} \qquad \textbf{(Equation 11)}$$

$$x4 - mx = \begin{bmatrix} 1/4 \\ -1/4 \\ 3/4 \end{bmatrix} \qquad \textbf{(Equation 12)}$$

These subtractions are used to calculate the product of each subtraction by itself, from the transposed matrix:

$$(x1 - mx)(x1 - mx)^T = \begin{pmatrix} 9/16 & 3/16 & 3/16 \\ 3/16 & 1/16 & 1/16 \\ 3/16 & 1/16 & 1/16 \end{pmatrix} \qquad \textbf{(Equation 13)}$$

$$(x2 - mx)(x2 - mx)^T = \begin{pmatrix} 1/16 & -1/16 & -1/16 \\ -1/16 & 1/16 & 1/16 \\ -1/16 & 1/16 & 1/16 \end{pmatrix} \qquad \textbf{(Equation 14)}$$

$$(x3 - mx)(x3 - mx)^T = \begin{pmatrix} 1/16 & 3/16 & -1/16 \\ 3/16 & 9/16 & -3/16 \\ -1/16 & -3/16 & 1/16 \end{pmatrix} \qquad \textbf{(Equation 15)}$$

$$(x4 - mx)(x4 - mx)^T = \begin{pmatrix} 1/16 & -1/16 & 3/16 \\ -1/16 & 1/16 & -3/16 \\ 3/16 & -3/16 & 9/16 \end{pmatrix} \qquad \textbf{(Equation 16)}$$

This sum is averaged, this result being the covariance matrix, as shown below:

$$(Cx) = \frac{1}{4}\begin{pmatrix} 12/16 & 4/16 & 4/16 \\ 4/16 & 12/16 & -4/16 \\ 4/16 & -4/16 & 12/16 \end{pmatrix}$$ **(Equation 17)**

or

$$(Cx) = \begin{pmatrix} 3/16 & 1/16 & 1/16 \\ 1/16 & 3/16 & -1/16 \\ 1/16 & -1/16 & 3/16 \end{pmatrix}$$ **(Equation 18)**

Considering a vector v that is an eigenvector of a square matrix M when we have the multiplication of M and v, resulting in a multiple of v, that is, the multiplication of a scalar λ by v.λ is then called the eigenvalue of the associated matrix M to the eigenvector v.

The desired property of eigenvectors is just the direction. So, when we say eigenvectors, we can understand it as “eigenvectors of length 1”, that is, not null. They are orthogonal to each other, so you can present the data using them instead of the x, y, z axes.

The characteristic equation of M was used to calculate the eigenvalues of 2x2 or 3x3 matrices:

$$det(M - \lambda \cdot I) = 0$$ **(Equation 19)**

where I is the identity matrix, M is the given matrix, and the non-zero scalars λ will be the eigenvalues, as we can see in the following example:

$$det = \begin{pmatrix} m11 - \lambda & m12 \\ m21 & m22 - \lambda \end{pmatrix} = 0$$ **(Equação 20)**

The result is an equation of the 2nd degree. In this way, we can calculate and substitute the roots in the following system, finding the eigenvectors that correspond to each eigenvalue:

$$\begin{pmatrix} m11 - \lambda & m12 \\ m21 & m22 - \lambda \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$ **(Equação 21)**

For larger dimensions, an iterative numerical algorithm is applied. As the last step, it sorts the eigenvalues in descending order and, consequently, associated eigenvectors. That is, PCs are constructed in descending order of the amount of variance they describe. For example, the first factor describes a greater variance of the data than the second, the second describes a greater variance than the third, etc.

You must also rearrange the eigenvectors in a decreasing order, according to the values of the associated eigenvalues. In this way, we have in the first position the eigenvector related to the largest eigenvalue, in the second position, the second eigenvalue, and so on. Considering ϵ as the eigenvectors, and λ as the eigenvalues, we will have ϵn-1 and λn-1.

Considering a matrix A, where the columns are the eigenvectors of Cx ordered in descending order, considering a transformation defined as:

$$y = A(x - mx)$$ **(Equation 22)**

This transformation will schematize the x values into y values, where the mean will be zero, that is, my = 0. The covariance matrix of the y of A and Cx can be obtained by:

$$Cx = A(CxA^T)$$ **(Equation 23)**

The matrix Cy has elements on the main diagonal that are the eigenvalues of Cx, we can then say that Cx and Cy have the same eigenvalues and eigenvectors (ANTON; RORRES, 2004). We get:

$$(Cy) = \begin{pmatrix} \lambda 1 & 0 & 0 \\ 0 & \lambda 2 & 0 \\ 0 & 0 & \lambda 3 \end{pmatrix} \qquad \textbf{(Equation 24)}$$

The elements that are not part of the main diagonal of Cy are zeros, so the elements of the vectors are uncorrelated.

So, the transformation mentioned at the beginning of this section: y=A(x-mx), is known as the Hotelling Transform, which is the alignment of the eigenvectors by rotation of the axis system, and this alignment is what de-correlates the Gonzalez and Woods data. (2000). It is possible to visualize this rotation of the axes according to Figure 2 (a, b, and c).

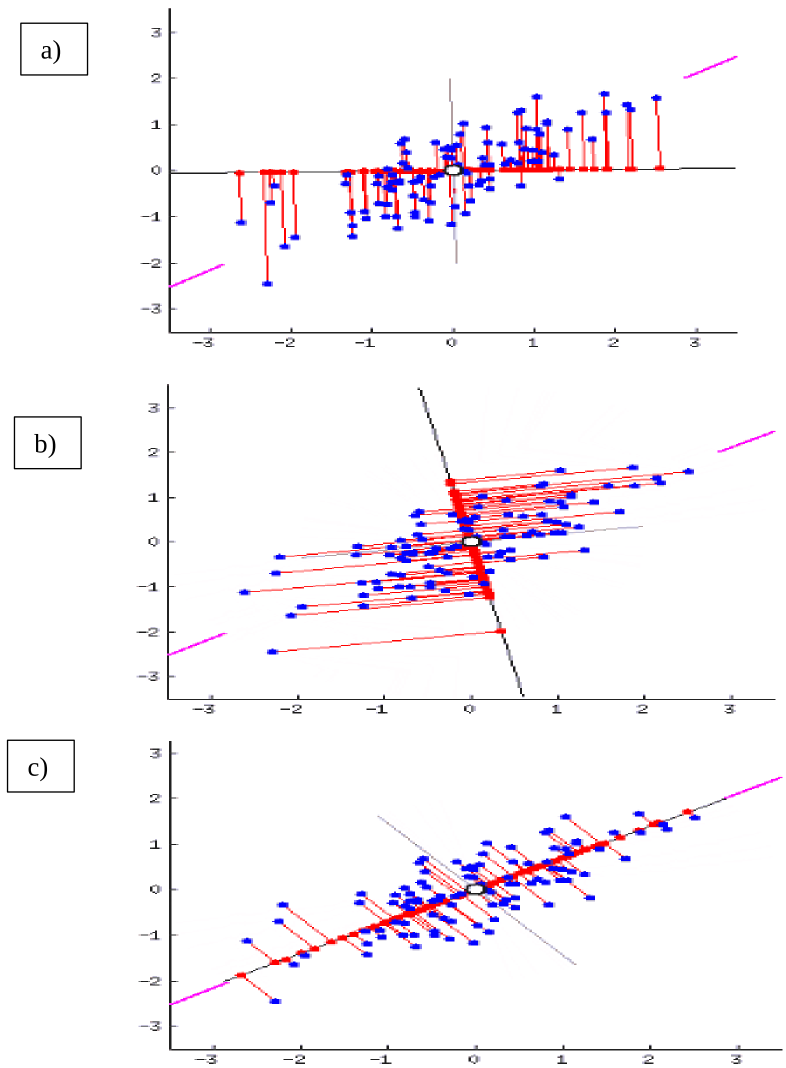


Figure 2 – Rotation of the main axes of any dataset.
Source: Modified from Pessoa (2019).

The blue dots in the figures are the data, the red lines are the distances, and the red dots are the variances. The distances should be minimal and the points in red more widely spaced (maximum variance). Figure 2 (c) shows the alignment that corresponds to the 1st Principal Component.

Each calculated eigenvalue shows the variance of the yi component along with the vector vi, this variance being an important aspect of the analysis of the input data.

2.3 Application of PCA to study data

To apply the PCA statistical method, the MATLAB Software was used. The MATLAB "pca" command allows the application of the "als.m" algorithm to work with missing data (Not a Number – NaN); a description of the application of PCA with LA to reduce information loss (Dray and Josse, 2015) is discussed in the Annex of da Silva (2021). The program returns all the variables necessary for evaluating the study data, these are listed below:

- ***coeff*:** Coefficients or eigenvectors of the principal components. Each coeff column contains coefficients for a PC. The columns are in descending order of variation.
- ***score*:** Scores of the main components. These are the projections on each principal component returned as an array. Score lines correspond to observations and columns to components.
- ***latent*:** Deviations from the principal components, that is, the eigenvalues. These are the eigenvalues of the covariance matrix of X.
- ***explained*:** Percentage of the total variation explained. Percentage of total change explained by each major component.
- ***mu*:** Estimated average. The estimated average of variables in X, is returned as a row vector when "Centered" is set to true. When "Centered" is false, the software returns a vector of zeros.

3. Results and discussion

The dendrochronological series were arranged in an excel file, in which the columns correspond to the samples, and the rows correspond to the years. This file is the input of the "pca" program. In this article, 37 PCs were obtained, corresponding to the 37 samples. Figure 3 shows the first four PCs that have the highest variances corresponding to the original series. The first PC explains 29.16% of the total variance and will be disregarded from the reconstruction of the dendrochronological series to remove the natural tendency for growth. In this sense, PCA is being used as a filter or as a model of the natural growth trend. It is worth mentioning that the remaining PCs (after the fourth) were omitted, as their variances are close to zero.

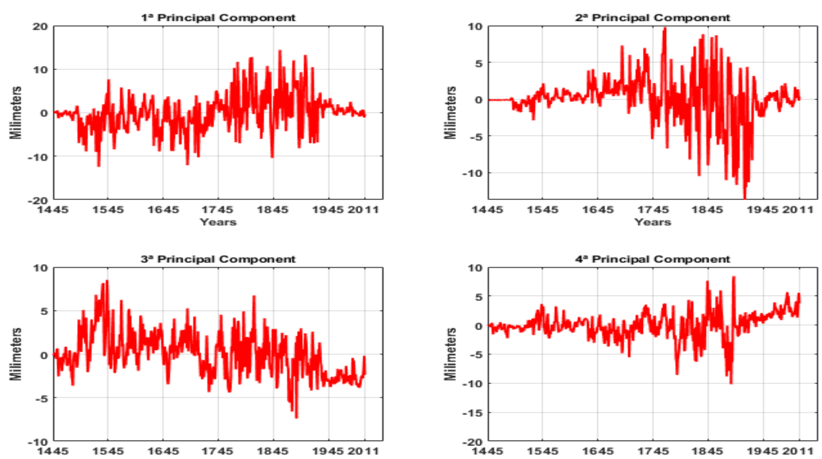


Figure 3 - Rotation of the principal axes of an example dataset.
Source: the authors (2020).

The variances of the first 24 PCs are represented in Table 3. The amount of information contained in a single principal component can be given by the percentage of explained variance. It is observed that more than 80% of the original information is contained in the first ten main components. The last 13 PCs were not shown because their values are too small, and these do not describe the set of samples well.

Table 3 – Percentage that each PC explains from the imbuia Dendrochronological Series. The PC column is the components, and the VA column is the percentage of variance.

PC	**VA**	**PC**	**VA**	**PC**	**VA**
1ª	21,01%	9ª	2,82%	17ª	1,12%
2ª	17,32%	10ª	2,61%	18ª	1,06%
3ª	10,48%	11ª	2,03%	19ª	0,95%
4ª	8,89%	12ª	1,77%	20ª	0,83%
5ª	6,26%	13ª	1,52%	21ª	0,79%
6ª	4,67%	14ª	1,41%	22ª	0,75%
7ª	3,92%	15ª	1,26%	23ª	0,67%
8ª	2,93%	16ª	1,24%	24ª	0,56%

Source: The authors (2020).

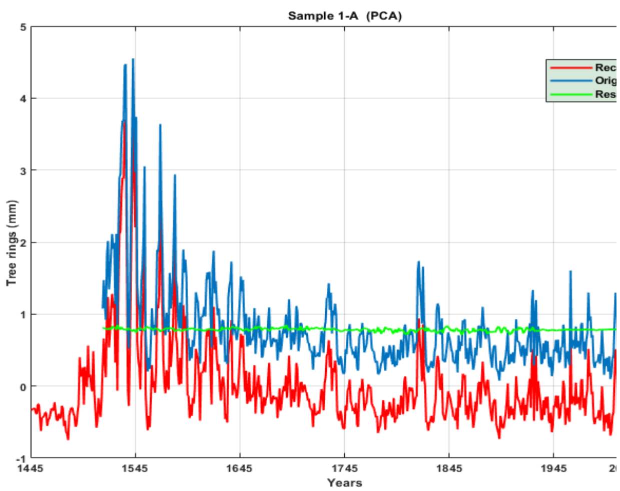
Then, the transformation to the original space was performed, that is, the reconstruction of the data, but without the 1st Principal Component (1st PC). It suggests that the 1st PC is representative of the natural growth trend (growth model) of all the trees in the place, for the species *Ocotea porosa (Nees & Mart) Barroso*, that is, the factor of internal influence on the development of trees. With this, we seek to maximize the signals of external factors (climatic and geophysical forcing) that acted in its development, improving dendroclimatic studies, and climatic reconstructions that can be carried out from these data.

Comparative graphs of the original series with the series filtered by PCA and the residue were constructed for all 37 samples. However, only three graphs were presented to visualize the results of the methodology of this article, Figure 4 (a, b, and c), as it is enough to understand the methodology and the results, considering that all the graphs obtained represent a dendrochronological series, the reconstructed series, and the residue.

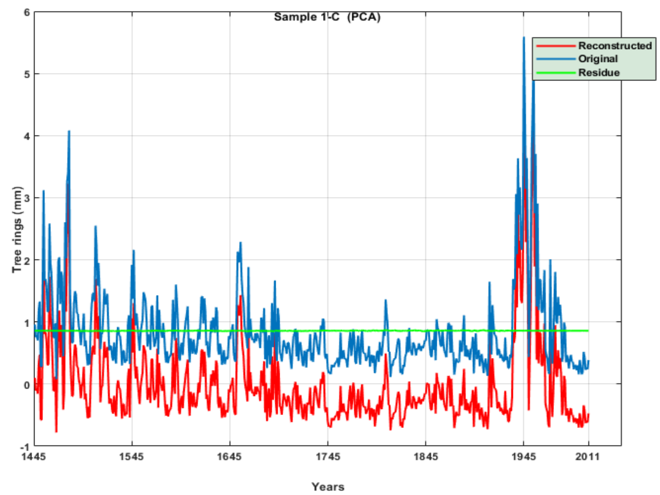
Figure 4 shows examples of reconstructed data without the 1st Principal Component (1st PC). In this figure, the original dendrochronological data corresponding to the sample (A, C, and B, respectively) referring to the trees referenced by us as "1" and "11" are presented. The reconstruction is shown in red, while the original data is shown in blue. The residual, which is the difference between these two time series, is shown in green.

a) Sample A from Tree 1.

a) Sample A from Tree 1.



b) Sample C from Tree 1.



c) Sample B from tree 11.

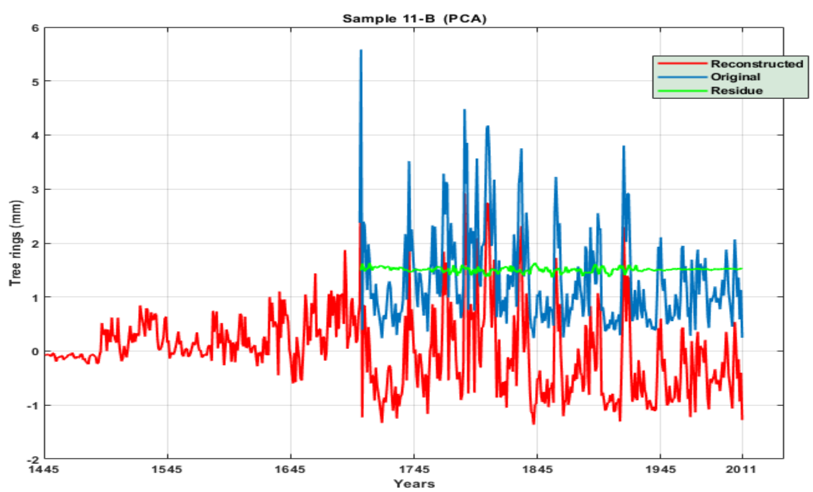


Figure 4 – Graphs of samples comparing the original dendrochronological series and the filtered dendrochronological series of each sample.
Source: the authors (2020).

An important detail to observe in the graphs of the series reconstructed by the PCA is that the method performs a reconstruction for the years before the birth of the smallest samples (those that are younger). This reconstruction is based on the other ray values of the time series, that is, based on all analyzed samples. The technique predicts how the youngest trees would grow if they had been born in the same year as the oldest tree. However, this computed period can be removed from the average chronology calculation. Citing an analogous case, in a study carried out by Silva et al. (2021), in which PCA was applied to remove natural trends, the period computed by PCA and ALS for the newer series was removed, for later comparison between other detrending methods. Therefore, comparing the present study with the study carried out by Silva et al. (2021), it is understood that this removal is at the discretion of the researcher, and may vary according to the type of objective that is intended to be achieved.

Another important observation is that PCA removes the high-frequency signals from the time series, which corresponds to the natural growth signals of trees. This maximizes low-frequency signals, which are characterized by external phenomena that influence tree growth. Thus, if the researcher's objective is to study, for example, the climate, this is the ideal method for detrending the samples. However, if the objective is to study tree growth and remove external signals, this is not the desired method. Information on high and low-frequency signals in dendrochronological series can be found in bibliographies specialized in tree ring studies, e.g., Speer (1971) and Cook (1990). The most common method for preserving low to medium frequency signals is Regional Curve Standardization (RCS) (Briffa and Melvin, 2011)

From the dendrochronological series filtered from each sample, the Dendrochronological Mean Series is calculated (Figure 5). After obtaining the average dendrochronological series filtered from internal influences, it can be used for further studies of climatic reconstructions. Filtering or "detrending" is a crucial step for dendroclimatic and paleoclimatic studies when the objective is to reconstruct past weather patterns. It is necessary to preserve the low-frequency signals when detrending, and in this way, maintain the weather signals that are present in the tree ring series ((HELAMA et al., 2004; HELAMA et al., 2017; MAES et al., 2017; ZHANG and CHEN, 2017). To this end, techniques such as Standardization of Regional Curves - RCS (BRIFFA and MELVIN, 2011) and Polynomial Fit - STD (COOK and PETERS, 1997) are widely used. Other methods, such as functions spline, end up removing some of these variances, and, consequently, are not recommended, as they can lead to the misinterpretation of the climate variability detected in the chronology of the tree rings. PCA is a statistical tool that allows keeping these low-frequency signals.

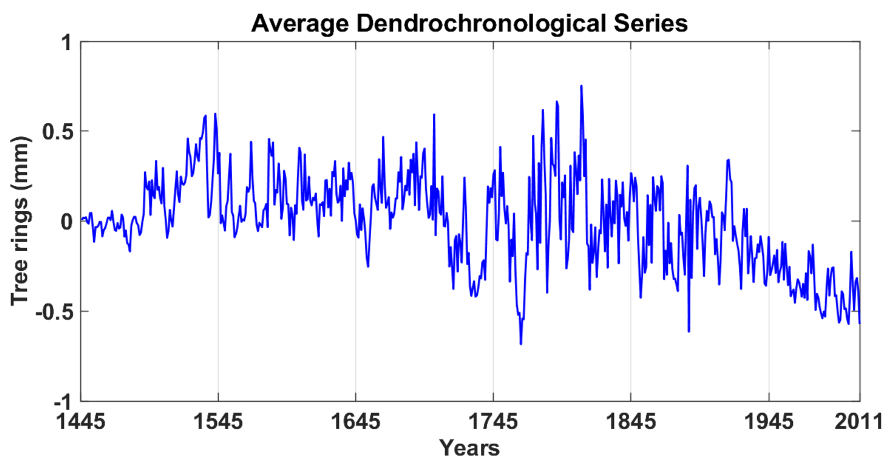


Figure 5 – Graph of the Dendrochronological Average Series obtained from the dendrochronological series of 37 samples.
Source The authors (2020).

4. Final considerations

A statistical analysis known as Principal Component Analysis (PCA) was performed to remove the natural growth trend of 37 samples of trees of the species *Ocotea porosa (Nees & Mart) Barroso* collected in the General Carneiro region.

PCA proved to be viable to obtain the regional growth trend for the species. The method makes a single representative fit to all samples collected, and can be used to remove more complicated biases, while other methods generally take on a specific function. Another very relevant result is the use of PCA in conjunction with ALS. With this, we were able to handle a series of tree rings with different lengths using an optimization algorithm. And consequently, there is no need to resize the series so that they have the same temporal length, which would cause data loss.

The best results were obtained for the longest temporal samples, i.e., the oldest tree samples. The use of the ALS optimization method may imply noise generation or peaks in the analysis of the youngest trees.

In addition to the aspects mentioned above, this work makes the application of the PCA method simpler and more didactic. In short, this method may be used to filter other tree samples, and to compose new dendrochronological studies.

Acknowledge

We thank CAPES for the master's scholarship, and for the PDSE scholarship (88881.624415/2021-01), CNPq process nº 305249/208-5, and FAPESP - 2009/02907-8 for the logistical structure of the laboratory.

References

Anderson, T. W. *An Introduction to Multivariate Statistical Analysis.* [S.l.]: John Wiley & Sons, Inc., 2003.

Anton, H.; Rorres, C. *Álgebra Linear com Aplicações*. Porto Alegre: Bookman, 2004.

Briffa, K.R.; Melvin, T. M. *A Closer Look at Regional Curve Standardization of Tree-Ring Records: Justification of the Need, a Warning of Some Pitfalls, and Suggested Improvements in Its Application*. Springer Netherlands, Dordrecht. pp. 113–145, 2011.

Cook, E.; Kairiukstis, L. *Methods of Dendrochronology: Applications in the Environmental Sciences.* Springer Netherlands, ISBN 9780792305866. Disponível em: https://books.google.com.br/books?id=zr8Ucld6FYcC, 1990.

Cook, E. R.; Peters, K., 1997. *Calculating unbiased tree-ring indices for the study of climatic and environmental change*. The Holocene, 7, 361–370, 1997.

DERGACHEV, V. A.; RASPOPOV, O. *The Long-Term Solar Cyclicity (210 and 90 years) and Variations of The Global Terrestrial Air Temperatures Since 1868*. In: *Solar & Space Weather Euroconference, The Solar Cycle and Terrestrial Climate*, v. 1, Santa Cruz de Tenerife: [s.n.], 2000.

Dray, S.; Josse, J. *Principal component analysis with missing values: a comparative survey of methods*. Plant Ecology, v. 216. doi: 10.1007/s11258-014-0406-z.

Fye, F.; Cleaveland, M. *Paleoclimatic analyses of tree-ring reconstructed summer drought in the United States, 1700-1978*. Tree-Ring Research, v. 57, p. 31–34, 2001.

Gonzalez, R.; Woods, R. *Processamento de Imagens Digitais*. Editora Blucher, 1 edição, 2000.

Helama, S.; Lindholm, M.; Timonen, M.; Eronen, M. *Detection of climate signal in dendrochronological data analysis: a comparison of tree-ring standardization methods*. Theoretical and Applied Climatology, 79. 2004.

Helama, S.; Melvin, T. M.; Briffa, K. R. *Regional curve standardization: State of the art.* The Holocene, 27, 172-177. 2017. doi: 10.1177/0959683616652709, 2016.

Hongyu, K.; Sandanielo, V. L. M.; Oliveira Junior, G. J. *Análise de componentes principais: Resumo teórico, aplicação e interpretação*. v. 5, p. 83–90, 2016.

LISI, C. S. *Atividade de 14C do Fallout e Razão Isotópica 13C/12C em Anéis de Crescimento de Clima Tropical e Subtropical do Brasil*. 2000. Tese (Doutorado) — Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, 2000.

Maes, S. L.; Vannoppen, A.; Altman, J.; Van Den Bulcke, J.; Decocq, G.; De Mil, T.; Depauw, L.; Landuyt.; Perring, M. P.; Van Acker, J.; Vanhellemont, M.; Verheyen, K. *Evaluating the robustness of three ring-width measurement methods for growth release reconstruction*. Dendrochronologia, 67-76. doi: https://doi.org/10.1016/j.dendro.2017.10.005, 2017.

PESSOA, W. *Reconhecimento de padrões - Eigenfaces.* Disponível em: https://medium.com/@williangp/reconhecimento-de-padr%C3%B5es-eigenfaces-e4cef8f04919. Acesso em 15/12/2019.

Raspopov, O.; Dergachev, V.; Kolström, T. *Periodicity of climate conditions and solar variability derived from dendrochronological and other paleoclimatic data in high latitudes*. Palaeogeography, Palaeoclimatology, Palaeoecology, v. 209, 2004.

Rigozo, N. R.; Nordemann, D. J. R.; Echer, E.; Antunes, L. E. *Search for solar periodicities in tree-ring widths from Concórdia (s.c., Brazil)*. Pure and Applied Geophysics, v. 161, p. 221–233, 2004.

Ricken, P.; Hess, A. F.; Borsoi, G. A. *Relações biométricas e ambientais no incremento diamétrico de Araucária angustifolia no Planalto Serrano Catarinense*. Ciência Florestal, v. 28, n. 4, p. 1592-1603, 2018.

Silva, D.O; Klausner, V.; Prestes, A.; Macedo, H. G.; Aakala, T.; Silva, I. R. *Principal components analysis: An alternative way for removing natural growth trends.* Pure and Applied Geophysics, https://doi.org/10.1007/s00024-021-02776-1, 2021.

Schweingruber, F. *Tree Rings: Basics and Applications of Dendrochronology*. [S.l.]: D. Reidel Publishing Company, 1988.

SHEPPARD, P.; GRAUMLICH, L. *A Reflected-Light Video Imaging System for Tree-Ring Analysis of Conifers.* In: *Proceedings of the 1994 International Conference on Tree Rings, Environment and Humanity*. Department of Geosciences, University of Arizona, Tucson: [s.n.], p. 879–889, 1996.

Speer, J. *Fundamentals of Tree-Ring Research*. [S.l.]: Library of Congress Cataloging-in-Publication Data, 1971.

Zhang, X.; Chen, Z. *A new method to remove the tree growth trend based on ensemble empirical mode decomposition.* Trees 31, 405–413, 2017.