



ISSN: 2447-3359

REVISTA DE GEOCIÊNCIAS DO NORDESTE

Northeast Geosciences Journal

v. 8, nº 1 (2022)

<https://doi.org/10.21680/2447-3359.2022v8n1ID24096>



Análise de componentes principais aplicada à dendrocronologia

Principal component analysis applied to dendrochronology

Daniela Oliveira da Silva¹; Virginia Klausner²; Alan Prestes³; Humberto Gimenes Macedo⁴

¹ Universidade do Vale do Paraíba, São José dos Campos/SP, Brasil. Email: fys.dani@gmail.com

ORCID: <https://orcid.org/0000-0002-3851-3368>

² Universidade do Vale do Paraíba, São José dos Campos/SP, Brasil. Email: viklausner@gmail.com

ORCID: <https://orcid.org/0000-0003-0250-5574>

³ Universidade do Vale do Paraíba, São José dos Campos/SP, Brasil. Email: aprestes@gmail.com

ORCID: <https://orcid.org/0000-0001-9403-8103>

⁴ Universidade do Vale do Paraíba, São José dos Campos/SP, Brasil. Email: gimenes Humberto@outlook.com

ORCID: <https://orcid.org/0000-0002-0858-6283>

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 (26°24'01 25"S 51°24'03 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.

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 (26°24'01 25"S 51°24'03 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.

1. Introdução

Dendrocronologia é o nome dado ao estudo dos anéis de crescimento das árvores. Esse tipo de técnica nos permite inferir como eram as condições ambientais na época de formação de cada anel. O pioneirismo nesta área é dado ao Astrônomo Andrew Ellicott Douglass, que notou, por volta do ano 1900, variações nas larguras dos anéis em algumas árvores. Essas variações foram atribuídas à influência de fatores externos (condições climáticas, atividade solar, nutrientes no solo, entre outros) e fatores internos (tendência natural de crescimento de determinada espécie). Portanto, as árvores são importantes registros de uma variedade de processos naturais do ambiente, agindo como um monitor de mudanças causadas pela humanidade, como, por exemplo, a poluição e a contaminação (SPEER, 1971).

Utiliza-se os dados dendrocronológicos para analisar os padrões espaciais e temporais da variação do clima diante de forçantes naturais (FYE; CLEAVELAND, 2001). Esse método consiste em determinar a idade das árvores analisando seus anéis formados em seus anos de crescimento, bem como, retirar outras informações, como a ocorrência de eventos climáticos (SCHWEINGRUBER, 1988). Esta técnica também é utilizada no estudo do ciclo das manchas solares. Diversos estudos observaram que as larguras dos anéis de crescimento de diferentes espécies, estavam relacionadas aos períodos de mínima e máxima atividade solar (DERGACHEV; RASPOPOV, 2000; RASPOPOV *et al.*, 2001; RIGOZO *et al.*, 2004).

Existem diversos métodos para análise de dados e a escolha do método depende das características dos mesmos e o que se deseja evidenciar pela análise em questão. Para dendrocronologia, comumente tem-se grande variedade de dados, ou seja, muitas amostras de árvores e, conseqüentemente, de raios medidos, com um grande número de anéis anuais. Sendo assim, faz-se necessário escolher uma análise estatística que auxilia na visualização desse grande volume de dados de uma forma simples, mas com a menor perda de informação possível.

Devido a estas peculiaridades, a Análise de Componentes Principais (do inglês Principal Component Analysis – PCA) demonstra ótima capacidade em trabalhar com séries dendrocronológicas. Segundo Anderson (2003), PCA consiste em combinações lineares de variáveis com propriedades especiais em relação às suas variâncias. Neste artigo, serão utilizadas séries temporais com diferentes tamanhos amostrais, portanto, se faz necessário trabalhar com o método dos mínimos quadrados alternados (do inglês Alternating Least Squares – ALS). A técnica ALS realiza uma extrapolação dos dados baseando-se nos originais. Isto é, as amostras mais velhas são utilizadas para estimar os valores faltantes e, desta forma, todas as amostras terão o mesmo tamanho. As componentes principais (PCs), calculadas pela PCA, facilitam a discriminação dos fatores internos e externos que atuaram no crescimento do grupo de árvores. Assim é possível avaliar fatores geofísicos e climáticos que influenciam no clima de determinada região e que ficaram registrados no desenvolvimento de cada anel de crescimento. Em outras palavras, a resposta de crescimento de cada espécie pode estar associada a fatores ambientais como temperatura, precipitação, umidade relativa, relação atmosfera-oceano etc., e, conseqüentemente, por meio destas respostas pode-se realizar reconstruções climáticas. Essas reconstruções irão depender da idade da cronologia da região de estudo.

O presente estudo tem por objetivo mostrar a vantagem da aplicação da metodologia de PCA na geração de série dendrocronológica de forma automatizada. A técnica tradicional de dendrocronologia realiza a remoção de sinais de influências internas ao crescimento da árvore de cada umas das amostras por meio iterativo, sendo mais custoso computacionalmente. A remoção da tendência de crescimento natural como um primeiro passo é fundamental para a análise de fatores externos, como, por exemplo, o estudo do clima da região de coleta das amostras de árvores. Além disso, este artigo apresenta o passo a passo matemático de um caso exemplo de aplicação da PCA, o que ajuda na compreensão do método.

2. Metodologia

2.1 Obtenção dos dados dendrocronológicos

Os dados dendrocronológicos são obtidos pela medição da espessura de cada anel de crescimento. Cada anel representa a variação de crescimento durante o período de um ano da árvore avaliada. Após realizadas todas as medições, obtém-se uma série temporal, i.e., a série dendrocronológica. A dendrocronologia média (Index) para a região de interesse do estudo é obtida por meio da média aritmética das larguras dos anéis das diversas séries dendrocronológicas (COOK;

KAIRIUKSTIS, 1990). Convém considerar para o cálculo do index o maior número de árvores possíveis dentro de uma dada região.

Os anéis de crescimento de algumas espécies de árvores apresentam áreas claras e escuras, como observado na Figura 1. As faixas claras representam a madeira formada no início do período de crescimento (lenho inicial), no qual têm paredes celulósicas finas e grandes diâmetros citoplasmáticos. As partes escuras estão relacionadas a madeira produzida no final do período de crescimento (lenho tardio), e têm paredes celulósicas mais espessas e menores diâmetros citoplasmáticos (LSI, 2000).

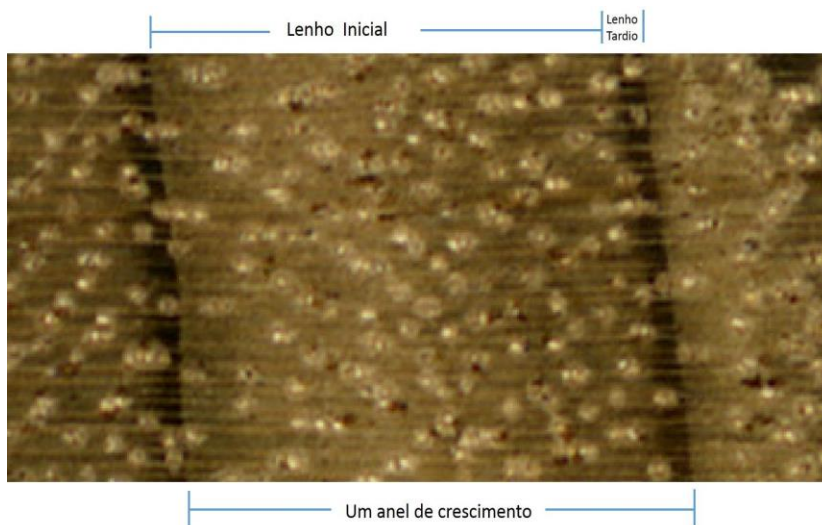
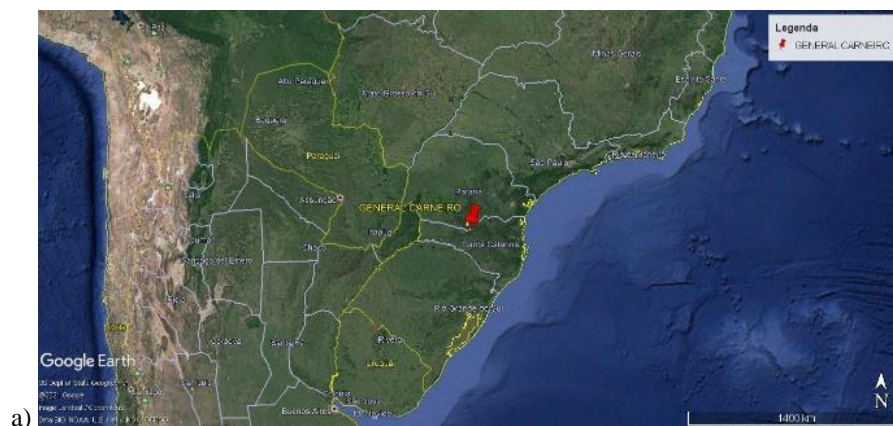


Figura 1 – Estrutura de um anel de crescimento da espécie *Ocotea porosa* (Nees & Mart) Barroso na qual é possível observar o lenho inicial e o lenho tardio.

Fonte: Autores (2020).

O conjunto de dados utilizado neste trabalho são séries de anéis de crescimento de árvores da espécie *Ocotea porosa* (Nees & Mart) Barroso, conhecida como Imbuia, provindas da região Sudeste do estado do Paraná, município de General Carneiro (Figuras 2a e 2b). Foram obtidas 64 amostras a partir de 21 árvores, coletadas em janeiro de 2013 por método não-destrutivo, com auxílio de sondas de Pressler, na altura do DAP (1,30 m). O número de amostras obtidas de cada árvore é apresentado na Tabela 1.



a)

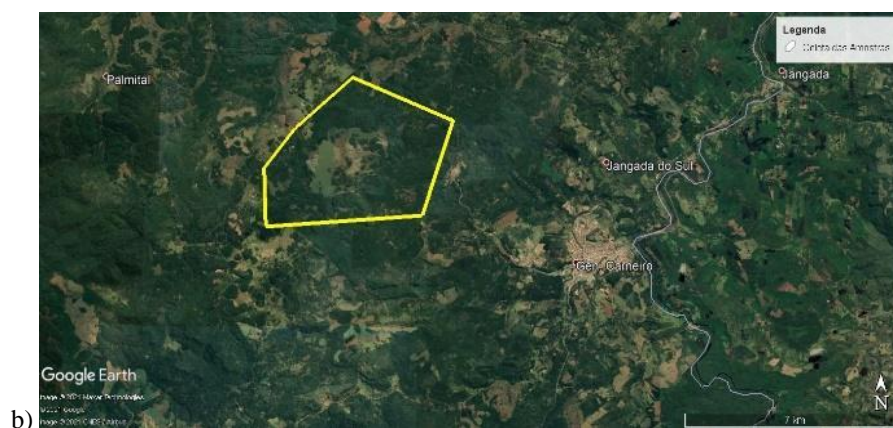


Figura 2 – Mapas da região de coleta. a) Mapa geral da região Sul do Brasil com a marcação da cidade de General Carneiro. b) Sítio de coleta das amostras na cidade de General Carneiro.

Fonte: Google Earth Pro (2020).

Tabela 1 – Número de amostras obtidas por árvore.

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			

Fonte: Autores (2020).

As amostras foram, inicialmente, polidas na superfície transversal com diferentes lixas (de 50 a 600 granos). Para realizar as medidas dos anéis de crescimento, utilizou-se uma mesa de mensuração VELMEX, com precisão de 0,001 mm, conectado a um microscópio estereoscópico com retículo micrométrico do Laboratório de Registros Naturais, Universidade do Vale do Paraíba - UNIVAP.

A datação foi feita no sentido da casca para a medula. O último anel foi formado até a data de coleta correspondente ao ano de 2011. Na região Sul brasileira, o início de formação de um anel de crescimento se dá na Primavera/Verão, e, desta forma, o anel correspondente ao ano de 2012 ainda não havia completado seu crescimento em janeiro de 2013 (período da coleta).

A escolha das amostras foi feita utilizando o processo hierárquico aglomerativo, com a medida de dissimilaridade entre os elementos (Análise de Clusters) dada pela distância Euclidiana Quadrática, empregando o método de variância de Ward (RICKEN *et al.*, 2018). Este procedimento é importante para que apenas árvores que tenham crescimento similar entrem no estudo, evitando assim aumentar o erro embutido. Então, das 64 amostras, foram escolhidas apenas 37, as quais estão listadas na Tabela 2.

Tabela 2 – Árvores e suas respectivas amostras selecionadas pela Análise de Clusters.

Árvore	Amostra (s)	Árvore	Amostra (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		

Fonte: Autores (2020).

2.2 Aplicação Matemática da PCA

Nesta seção será tratada a aplicação matemática da PCA com um exemplo numérico simples.

As PCs são obtidas com a diagonalização de matrizes simétricas positivas semi-definidas. Existem diversos programas com capacidade para cálculos matriciais como a diagonalização (HONGYU; SANDANIELO; OLIVEIRA JUNIOR, 2016). O exemplo numérico fornecido por Gonzalez e Woods (2000), e apresentado a seguir, ilustra a obtenção das PCs.

A covariância (ou variância) representa o grau com que cada variável é linearmente correlacionada com a outra. A equação para o cálculo de dados de duas dimensões é:

$$cov(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n} \quad \text{(Equação 1)}$$

sendo x e y listas de dados, com x sendo a primeira dimensão e y sendo a segunda dimensão; e são as médias das listas; x_i e y_i são cada um dos elementos das listas, na i -ésima posição; n é o número de itens dos dados. Quando os dados iniciam no índice 0 (representando uma amostra), no denominador e no somatório usa-se $n-1$, caso seja o conjunto total, usa-se o próprio n . Para dados com mais de duas dimensões, faz-se a covariância (Equação 1) a cada par de dimensões. Caso seja uma matriz de três dimensões (x , y e z), temos a seguinte matriz de covariância:

$$matrizcov = \begin{pmatrix} cov(x, x) & cov(x, y) & cov(x, z) \\ cov(y, x) & cov(y, y) & cov(y, z) \\ cov(z, x) & cov(z, y) & cov(z, z) \end{pmatrix} \quad \text{(Equação 2)}$$

Para dados dendrocronológicos, cada dimensão da matriz é dada por uma amostra de uma árvore diferente. Da diagonal principal desta matriz obtêm-se as variâncias, e nas outras posições têm-se as correlações entre as direções. Como esta matriz é simétrica e real, sempre pode-se encontrar um conjunto de autovetores ortonormais (ANTON; RORRES, 2004) que são os eixos principais da matriz.

Considerando M amostras de vetores, pode-se calcular o vetor médio (m_x) por:

$$mx = \frac{1}{M} \sum_{i=1}^M xi \quad \text{(Equação 3)}$$

Para exemplificar, usaremos um conjunto de vetores 3D com 4 amostras para cada passo destes cálculos. Supondo os seguintes valores para cada uma das 4 amostras:

$$x1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{(Equação 4)}$$

$$x2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{(Equação 5)}$$

$$x3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \text{(Equação 6)}$$

$$x4 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \quad \text{(Equação 7)}$$

Teremos o seguinte vetor médio:

$$mx = \begin{bmatrix} 3/4 \\ 1/4 \\ 1/4 \end{bmatrix} \quad \text{(Equação 8)}$$

Faz-se a subtração de cada xi do vetor médio (mx) para o cálculo da matriz de covariância:

$$x1 - mx = \begin{bmatrix} -3/4 \\ -1/4 \\ -1/4 \end{bmatrix} \quad \text{(Equação 9)}$$

$$x2 - mx = \begin{bmatrix} 1/4 \\ -1/4 \\ -1/4 \end{bmatrix} \quad \text{(Equação 10)}$$

$$x3 - mx = \begin{bmatrix} 1/4 \\ 3/4 \\ -1/4 \end{bmatrix} \quad \text{(Equação 11)}$$

$$x4 - mx = \begin{bmatrix} 1/4 \\ -1/4 \\ 3/4 \end{bmatrix} \quad \text{(Equação 12)}$$

Utiliza-se essas subtrações para calcular o produto de cada subtração por ela mesma, a partir da matriz transposta:

$$(x1 - mx)(x1 - mx)^T = \begin{pmatrix} 9/16 & 3/16 & 3/16 \\ 3/16 & 1/16 & 1/16 \\ 3/16 & 1/16 & 1/16 \end{pmatrix} \quad \text{(Equação 13)}$$

$$(x2 - mx)(x2 - mx)^T = \begin{pmatrix} 1/16 & -1/16 & -1/16 \\ -1/16 & 1/16 & 1/16 \\ -1/16 & 1/16 & 1/16 \end{pmatrix} \quad \text{(Equação 14)}$$

$$(x3 - mx)(x3 - mx)^T = \begin{pmatrix} 1/16 & 3/16 & -1/16 \\ 3/16 & 9/16 & -3/16 \\ -1/16 & -3/16 & 1/16 \end{pmatrix} \quad \text{(Equação 15)}$$

$$(x4 - mx)(x4 - mx)^T = \begin{pmatrix} 1/16 & -1/16 & 3/16 \\ -1/16 & 1/16 & -3/16 \\ 3/16 & -3/16 & 9/16 \end{pmatrix} \quad \text{(Equação 16)}$$

Faz-se a média desta soma, sendo este resultado a matriz de covariância, como é mostrado a seguir:

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

ou

$$(Cx) = \begin{pmatrix} 3/16 & 1/16 & 1/16 \\ 1/16 & 3/16 & -1/16 \\ 1/16 & -1/16 & 3/16 \end{pmatrix} \quad \text{(Equação 18)}$$

Temos que um vetor v é um autovetor de uma matriz quadrada M quando temos a multiplicação de M e v , resultando em um múltiplo de v , isto é, na multiplicação de um escalar λ por v . λ é então chamado de autovalor da matriz M associado ao autovetor v .

A propriedade desejada dos autovetores é apenas a direção. Então quando se diz autovetores, pode-se entender como “autovetores de comprimento 1”, ou seja, não nulos. Eles são ortogonais entre si, então se pode apresentar os dados utilizando-os no lugar dos eixos x , y , z .

Utilizou-se a equação característica de M para calcular os autovalores de matrizes 2x2 ou 3x3:

$$\det(M - \lambda \cdot I) = 0 \quad \text{(Equação 19)}$$

sendo I a matriz identidade, M a matriz cedida e os escalares não nulos λ serão os autovalores, podemos visualizar no exemplo a seguir:

$$\det = \begin{pmatrix} m_{11} - \lambda & m_{12} \\ m_{21} & m_{22} - \lambda \end{pmatrix} = 0 \quad \text{(Equação 20)}$$

Têm-se como resultado uma equação do 2º grau. Deste modo, pode-se calcular e substituir as raízes no sistema a seguir, encontrando os autovetores que correspondem a cada autovalor:

$$\begin{pmatrix} m_{11} - \lambda & m_{12} \\ m_{21} & m_{22} - \lambda \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{(Equação 21)}$$

Para dimensões maiores, é aplicado um algoritmo numérico iterativo. Como último passo, ordena de forma decrescente os autovalores, e, conseqüentemente, autovetores associados. Ou seja, as PCs são construídas em ordem decrescente da quantidade de variância que descrevem. Por exemplo, o primeiro fator descreve maior variância dos dados que o segundo, o segundo descreve maior variância que o terceiro, etc.

Deve-se também rearranjar os autovetores de forma decrescente, de acordo com os valores dos autovalores associados. Desta forma, temos na primeira posição o autovetor relacionado com o maior autovalor, na segunda posição, o segundo autovalor, e assim por diante. Considerando e como os autovetores, e λ como os autovalores, teremos $e\lambda - I$ e $\lambda n - I$.

Considerando uma matriz A , onde as colunas são os autovetores de Cx ordenados de forma decrescente, considerando uma transformação definida como:

$$y = A(x - mx) \quad \text{(Equação 22)}$$

Esta transformação irá esquematizar os valores x em valores y , onde a média será zero, ou seja, $my = 0$. Poderá se obter a matriz de covariância dos y de A e Cx por:

$$Cx = A(CxA^T) \quad \text{(Equação 23)}$$

A matriz Cy tem elementos na diagonal principal que são os autovalores de Cx , podemos afirmar então que Cx e Cy possuem os mesmo autovalores e autovetores (ANTON; RORRES, 2004). Obtêm-se:

$$(Cy) = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \quad \text{(Equação 24)}$$

Os elementos que não fazem parte da diagonal principal de Cy são zeros, então os elementos dos vetores são descorrelacionados.

Então, a transformação dita no início desta seção: $y=A(x-mx)$, é conhecida como a Transformada de *Hotelling*, que é o alinhamento dos autovetores por rotação do sistema de eixos, sendo este alinhamento quem descorrelaciona os dados Gonzalez e Woods (2000). É possível visualizar essa rotação dos eixos segundo Figura 2 (a, b e c).a)

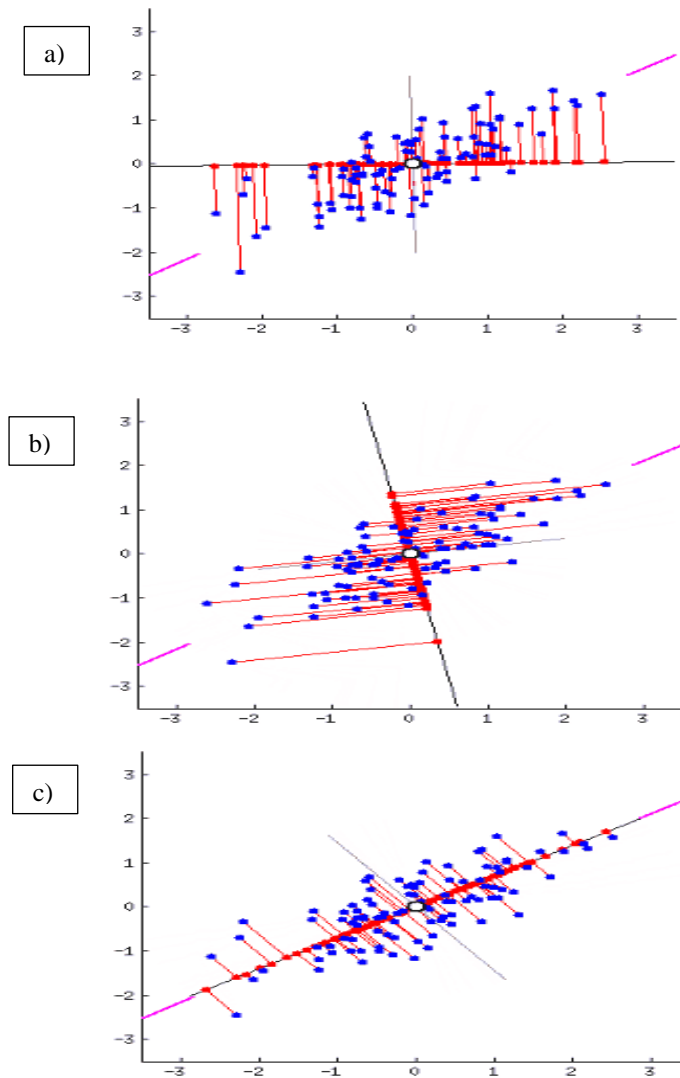


Figura 2 – Rotação dos eixos principais de um conjunto de dados qualquer.
Fonte: Modificada de Pessoa (2019).

Os pontos azuis das figuras são os dados, as linhas em vermelho são as distâncias, e os pontos em vermelho as variâncias. É desejável que as distâncias sejam mínimas e os pontos em vermelho mais espaçados (máxima variância). A Figura 2 (c) mostra o alinhamento que corresponde a 1ª Componente Principal.

Cada autovalor calculado mostra a variância do componente y_i ao longo do vetor v_i , sendo esta variância um aspecto importante para a análise dos dados de entrada.

2.3 Aplicação da PCA aos dados do estudo

Para aplicar o método estatístico PCA, foi utilizado o Software MATLAB. O comando “pca” do MATLAB possibilita a aplicação do algoritmo “als.m” para se trabalhar com dados ausentes (do inglês *Not a Number* – NaN); uma descrição da aplicação da PCA com o ALS, para reduzir a perda de informações (Dray and Josse, 2014) é discutida no Anexo de Silva (2021). O programa retorna todas as variáveis necessárias para avaliação dos dados de estudo, estas estão relacionadas a seguir:

- **coeff**: Coeficientes ou autovetores das componentes principais. Cada coluna do *coeff* contém coeficientes para uma PC. As colunas estão na ordem de variação descendente.
- **score**: Pontuações dos componentes principais. São as projeções em cada componente principal retornadas como uma matriz. As linhas de pontuação correspondem a observações e colunas a componentes.
- **latent**: Desvios das componentes principais, ou seja, os autovalores. São os valores próprios da matriz de covariância de X.
- **explained**: Porcentagem da variação total explicada. Porcentagem da variação total explicada por cada componente principal.
- **mu**: Média estimada. Média estimada das variáveis em X, retornada como um vetor de linha quando “Centered” está definido como verdadeiro. Quando “Centered” é falso, o software retorna um vetor de zeros.

3. Resultados e discussão

As séries dendrocronológicas foram dispostas em um arquivo excel, no qual as colunas correspondem às amostras, e as linhas correspondem aos anos. Esse arquivo é a entrada do programa “pca”. Neste artigo, foram obtidas 37 PCs, correspondentes às 37 amostras. Na Figura 3 são apresentadas as quatro primeiras PCs que tem as maiores variâncias correspondentes à série original. A primeira PC explica 29,16% do total de variância, e será desconsiderada da reconstrução das séries dendrocronológicas a fim de remover a tendência natural de crescimento. Nesse sentido, a PCA está sendo utilizada como filtro ou como um modelo de tendência natural de crescimento. Vale ressaltar que as PCs restantes (após a quarta) foram omitidas, pois suas variâncias são próximas de zero.

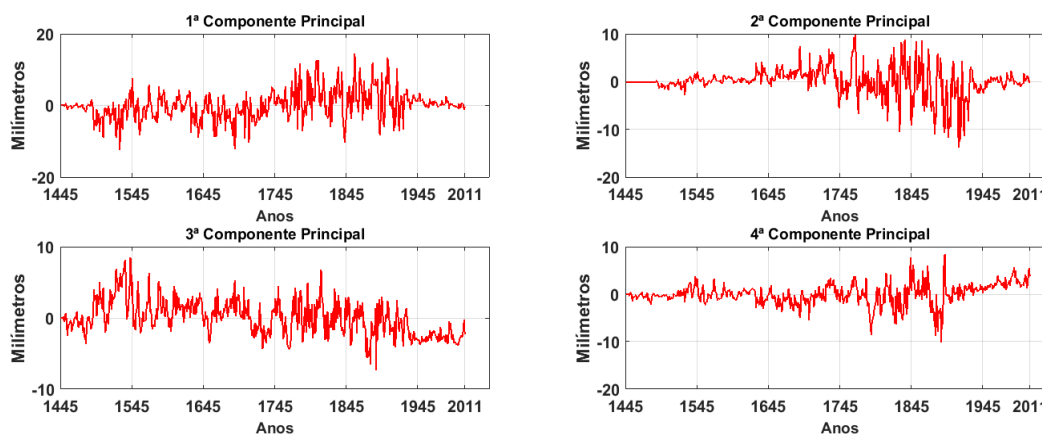


Figura 3 – Rotação dos eixos principais de um conjunto de dados exemplo.
Fonte: os autores (2020).

As variâncias das primeiras 24 PCs estão representadas na Tabela 3. A quantidade de informação contida numa única componente principal pode ser dada pela percentagem de variância explicada. Observa-se que mais de 80% da informação original está contida nas dez primeiras componentes principais. As últimas 13 PCs não foram mostradas pois seus valores são muito pequenos, e estas não descrevem bem o conjunto das amostras.

Tabela 3 – Percentagem que cada PC explica da Série Dendrocronológica de imbuia. A coluna PC são as componentes e a coluna VA a percentagem de variância.

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

Fonte: os autores (2020).

Em seguida, foi realizada a transformação para o espaço original, ou seja, a reconstrução dos dados, mas sem a 1^a Componente Principal (1^a PC). Sugere que a 1^a PC é representativa da tendência de crescimento natural (modelo de crescimento) de todas as árvores do local, para a espécie *Ocotea porosa* (Nees & Mart) Barroso, ou seja, o fator de influência interna no desenvolvimento das árvores. Com isso, procura-se maximizar os sinais dos fatores externos (forçantes climáticos e geofísicos) que atuaram em seu desenvolvimento, melhorando estudos dendroclimáticos, e a reconstruções climáticas que podem ser realizadas a partir destes dados.

Foram construídos gráficos comparativos da série original com a série filtrada pela PCA e o resíduo, para todas as 37 amostras. Porém, foram apresentados apenas três gráficos para visualização dos resultados da metodologia deste artigo, Figura 4 (a, b e c), pois é o suficiente para entender a metodologia e os resultados, considerando que todos os gráficos obtidos representam uma série dendrocronológica, a série reconstruída e o resíduo.

A Figura 4 mostra exemplos de dados reconstruídos sem a 1^a Componente Principal (1^a PC). Nesta figura, são apresentados os dados dendrocronológicos originais correspondentes à amostra (A, C, e B, respectivamente) referente às árvores referenciadas por nós como “1” e “11”. A reconstrução é mostrada em vermelho, enquanto os dados originais são mostrados em azul. O resíduo, que é a diferença entre essas duas séries temporais, é mostrado em verde.

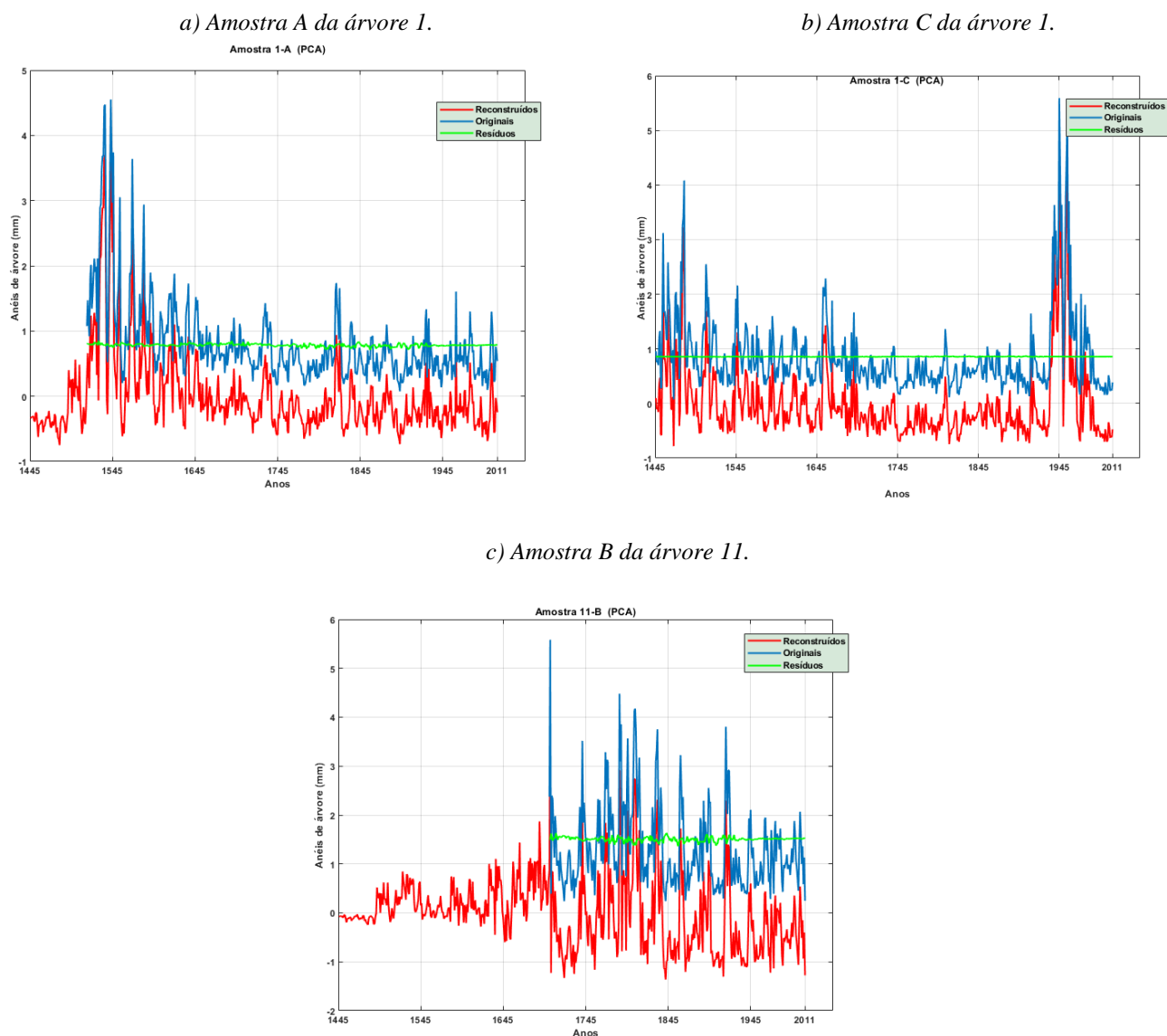


Figura 4 – Gráficos das amostras comparando a série dendrocronológica original e a série dendrocronológica filtrada de cada amostra.

Fonte: Os autores (2020).

Um detalhe importante para se observar nos gráficos das séries reconstruídas pela PCA, é que o método realiza uma reconstrução para os anos anteriores ao nascimento das menores amostras (as que são mais jovens). Essa reconstrução é baseada nos demais valores de raios das séries temporais, ou seja, baseado em todas as amostras analisadas. A técnica prevê como seria o crescimento das árvores mais jovens se estas tivessem nascido no mesmo ano da árvore mais velha. Entretanto, este período computado pode ser removido do cálculo da cronologia média. Citando caso análogo, em um estudo realizado por Silva *et al.* (2021), no qual a PCA foi aplicada para remoção das tendências naturais, o período computado pela PCA e ALS para as séries mais novas foi removido, para posterior comparação entre outros métodos de *detrending*. Portanto, comparando o presente estudo com o estudo realizado por Silva *et al.* (2021), entende-se que esta remoção fica a critério do pesquisador, e pode variar de acordo com o tipo objetivo que se pretende alcançar.

Outra observação importante, a PCA retira os sinais de alta frequência da série temporal, isto é, correspondente aos sinais de crescimento natural das árvores. Assim, maximiza os sinais de baixa frequência. Esses sinais são caracterizados pelos fenômenos externos que influenciam no crescimento das árvores. Desta forma, se o objetivo do pesquisador é estudar, por exemplo, o clima, este será o método ideal para *detrending* das amostras. Entretanto, caso o objetivo seja estudar o crescimento das árvores e remover sinais externos, este não será um método desejado. Informações quanto a sinais de alta e baixa frequência em séries dendrocronológicas podem ser encontrados em bibliografias especializadas em estudos dos anéis de árvores, e.g., Speer (1971) e Cook e Kairiukstis (1990). O método mais comum para preservar sinais de baixa a média frequência é a *Regional Curve Standardization* (RCS) (Briffa e Melvin, 2011).

A partir das séries dendrocronológicas filtradas de cada amostra, calcula-se a Série Média Dendrocronológica (Figura 5). Obtida a série média dendrocronológica filtrada de influências internas, pode-se utilizá-la para estudos posteriores de reconstruções climáticas. A filtragem, ou “*detrending*” é um passo crucial para estudos dendroclimáticos e paleoclimáticos quando o objetivo é reconstruir padrões climáticos do passado. É necessário preservar os sinais de baixa frequência ao fazer o *detrending*, e deste modo, manter os sinais climáticos que estão presentes na série de anéis de árvores (HELAMA et al., 2004; HELAMA et al., 2017; MAES et al., 2017; ZHANG e CHEN, 2017). Para este fim, técnicas como Padronização de Curvas Regionais – RCS (BRIFFA e MELVIN, 2011) e Ajuste Polinomial – STD (COOK e PETERS, 1997) são bastante utilizadas. Outros métodos, como funções *spline*, acabam removendo algumas dessas variâncias, e, conseqüentemente, não são recomendados, pois podem levar à interpretação errônea da variabilidade do clima detectada na cronologia dos anéis das árvores. A PCA é uma ferramenta estatística que permite manter estes sinais de baixa frequência.

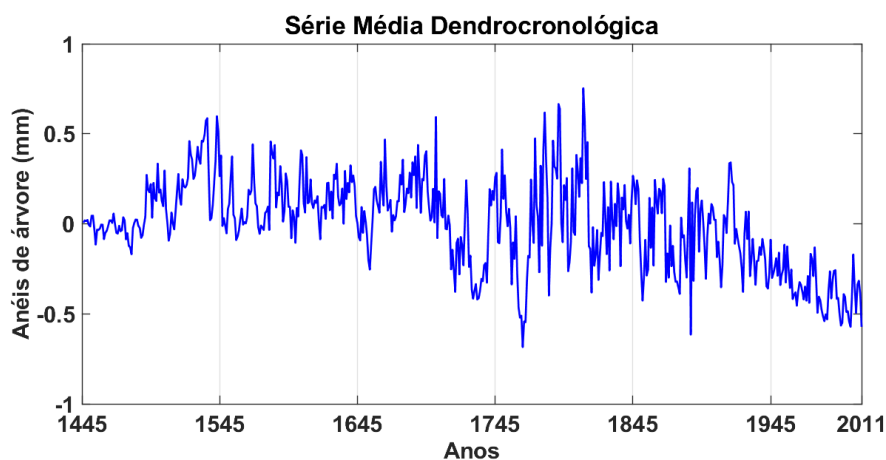


Figura 5 – Gráfico da Série Média Dendrocronológica obtida a partir das séries dendrocronológicas de 37 amostras.
Fonte: Os autores (2020).

4. Considerações finais

Realizou-se uma análise estatística conhecida como Análise de Componentes Principais (PCA) para remoção da tendência de crescimento natural das 37 amostras de árvores da espécie *Ocotea porosa* (Nees & Mart) Barroso coletadas na região de General Carneiro.

A PCA se mostrou viável para obter a tendência de crescimento regional para a espécie. O método faz um único ajuste representativo a todas as amostras coletadas, e pode ser usado para remover tendência mais complicadas, enquanto outros métodos geralmente assumem uma função específica. Outro resultado bastante relevante, é o emprego da PCA em conjunto com o ALS. Com isso, nós conseguimos lidar com séries de anéis de árvore com diferentes comprimentos usando um algoritmo de otimização. E conseqüentemente, não há necessidade de redimensionar as séries para que as mesmas tenham o mesmo comprimento temporal, o que causaria perda de dados.

Os melhores resultados foram obtidos para as amostras temporais mais longas, ou seja, as amostras de árvores mais velhas. A utilização do método de otimização ALS pode implicar em geração de ruído ou picos na análise das árvores mais jovens.

Além dos aspectos mencionados acima, este trabalho torna mais simples e didática a aplicação do método PCA. E, em um futuro próximo, este método poderá ser utilizado na filtragem de outras amostras de árvores, e para compor novos estudos dendrocronológicos.

Agradecimentos

Agradecemos a CAPES pela bolsa de mestrado, e pela bolsa PDSE (88881.624415/2021-01), ao CNPq processo nº 305249/208-5, e a FAPESP - 2009/02907-8 pela estrutura logística do laboratório.

Referências

- 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 Cyclicality (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.; Vanhellefont, 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.