SSA (Singular Spectrum Analysis)

André Almeida de Mello

Observações Preliminares

Decomposição espectral de uma matriz quadrada

$$A = P\Lambda P^{-1}$$

$$A = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^T + \lambda_2 \mathbf{v}_2 \mathbf{v}_2^T + \ldots + \lambda_n \mathbf{v}_n \mathbf{v}_n^T$$

Decomposição em valores singulares de uma matriz retangular

$$A = U\Sigma V^T$$

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \ldots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$$

Introdução

A SSA

- É um método não paramétrico para análise e previsão de séries temporais.
- Consiste basicamente na decomposição aditiva da série temporal em componentes de tendência, periódicos (ou harmônicos) e ruído, e em sua posterior reconstrução.
- Pode ser empregada, por exemplo, na identificação de componentes sazonais de diferentes períodos, na realização de previsões e na filtragem de ruído.

Vantagens e desvantagens

- Vantagens: não faz suposições sobre a forma funcional da série temporal, além de poder ser aplicada a séries temporais não estacionárias.
- Limitações: sensível à escolha de parâmetros, como o comprimento das trajetórias, além de ser computacionalmente intensiva para séries temporais muito longas.

Visão geral do algoritmo SSA

- O algoritmo SSA pode ser dividido em dois estágios: I) decomposição e II) reconstrução.
- Por sua vez, cada estágio pode ser dividido em dois passos.
- O estágio da decomposição se divide em 1) incorporação (*embedding)* e 2) decomposição SVD.
- O estágio da reconstrução se divide em 3) agrupamento e 4) médias diagonais.

Estágio I: decomposição

Passo 1: Incorporação

A série temporal original é mapeada em uma sequência de vetores defasados, constituindo a chamada “matriz trajetória”:

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_K \\ x_2 & x_3 & x_4 & \cdots & x_{K+1} \\ x_3 & x_4 & x_5 & \cdots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \cdots & x_N \end{bmatrix}$$

Passo 2: decomposição em valores singulares

Este passo consiste simplesmente na decomposição em valores singulares da matriz trajetória da série temporal em análise:

$$\mathbf{X} = \sum_{i=1}^{L} = P_i Q_i^T = \sum_{i=1}^{L} \sqrt{\lambda_i} U_i V_i^T$$

Estágio II: reconstrução

Passo 3: agrupamento

- Após o passo 2, a matriz trajetória terá sido decomposta como uma soma de L matrizes elementares.
- Este passo consiste simplesmente em agrupar tais matrizes em m grupos, conforme sejam identificadas como elementos da tendência, sazonalidade ou ruído.
- Temos agora:

$$\mathbf{X} = \mathbf{X}_{\mathbf{I_1}}, + \cdots + \mathbf{X}_{\mathbf{I_m}},$$

Passo 4: média diagonal

- Neste passo m séries temporais são obtidas das matrizes referentes a cada um dos grupos do passo anterior.
- Isto é feito por meio do cálculo da média aritmética das antidiagonais das matrizes, o que resulta em matrizes de Hankel, com estrutura similar à matriz trajetória.
- De cada uma destas matrizes de Hankel é recuperada uma série temporal que é um componente aditivo da série original.

Um exemplo simples

Passo 1: incorporação

$$\{x_t\} = \{1, 2, 3, 4, 5, 6\}, \; t = 1, \ldots, 6.$$

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \end{bmatrix}$$

Passo 2: decomposição em valores singulares

$$\mathbf{XX^T} = \begin{bmatrix} 30 & 40 & 50 \\ 40 & 54 & 68 \\ 50 & 68 & 86 \end{bmatrix}$$

$$\lambda_1 \approx 169.29$$

$$\lambda_2 \approx 0.71$$

$$\lambda_3 \approx 0$$

Passo 2: decomposição em valores singulares

$$U_1 = \begin{bmatrix} -0.418 \\ -0.565 \\ -0.712 \end{bmatrix}$$

$$U_2 = \begin{bmatrix} 0.812 \\ 0.12 \\ -0.572 \end{bmatrix}$$

$$U_3 = \begin{bmatrix} 0.408 \\ -0.816 \\ 0.408 \end{bmatrix}$$

$$V_1 = \begin{bmatrix} -0.283 \\ -0.413 \\ -0.543 \\ -0.674 \end{bmatrix}$$

$$V_2 = \begin{bmatrix} -0.788 \\ -0.361 \\ -0.066 \\ -0.494 \end{bmatrix}$$

Passos 3: agrupamento

$$\mathbf{X} = \mathbf{X_1} + \mathbf{X_2}$$

$$\mathbf{X_1} = \sqrt{\lambda_1} U_1 V_1^T = \begin{bmatrix} 1.539 & 2.247 & 2.953 & 3.661 \\ 2.079 & 3.036 & 3.993 & 4.950 \\ 2.261 & 3.827 & 5.033 & 6.239 \end{bmatrix}$$

$$\mathbf{X_2} = \sqrt{\lambda_2} U_2 V_2^T = \begin{bmatrix} -0.539 & -0.247 & 0.047 & 0.339 \\ -0.079 & -0.036 & 0.007 & 0.050 \\ 0.379 & 0.173 & -0.033 & -0.239 \end{bmatrix}$$

Aplicação no R

Série CO2

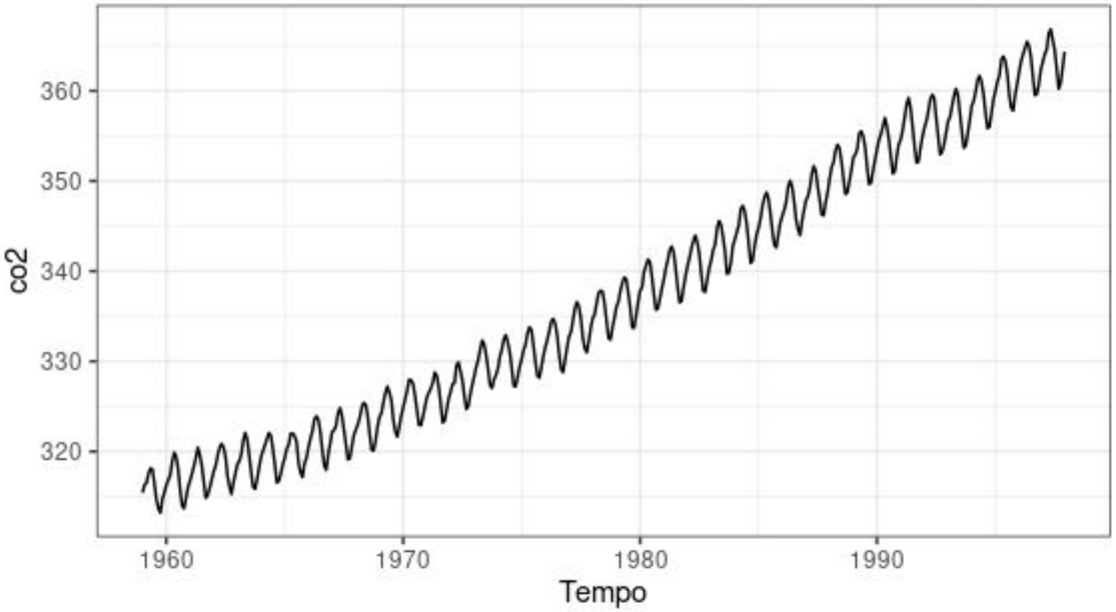


Gráfico dos autovalores obtidos

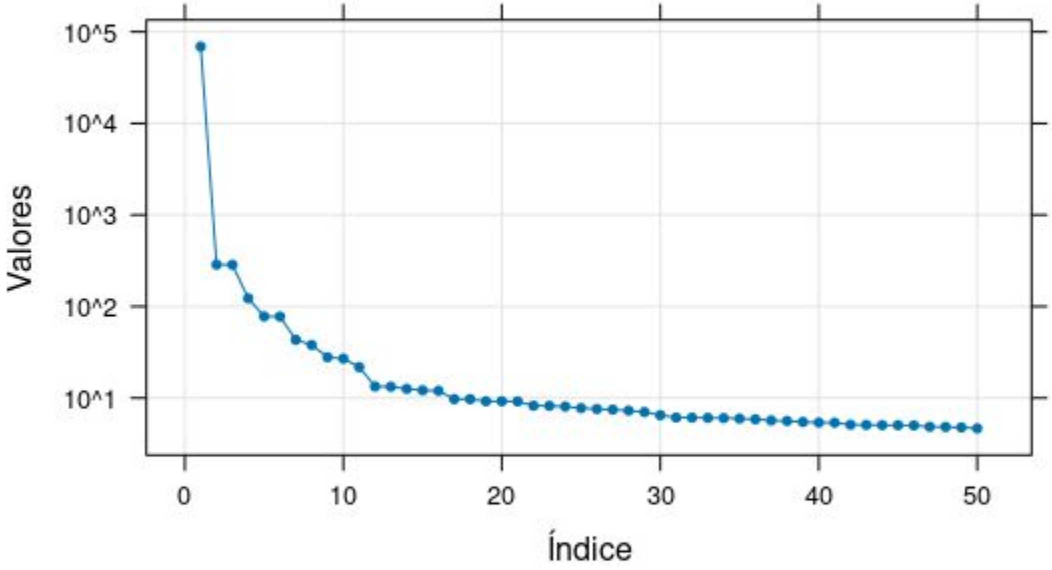


Gráfico dos dez primeiros autovetores

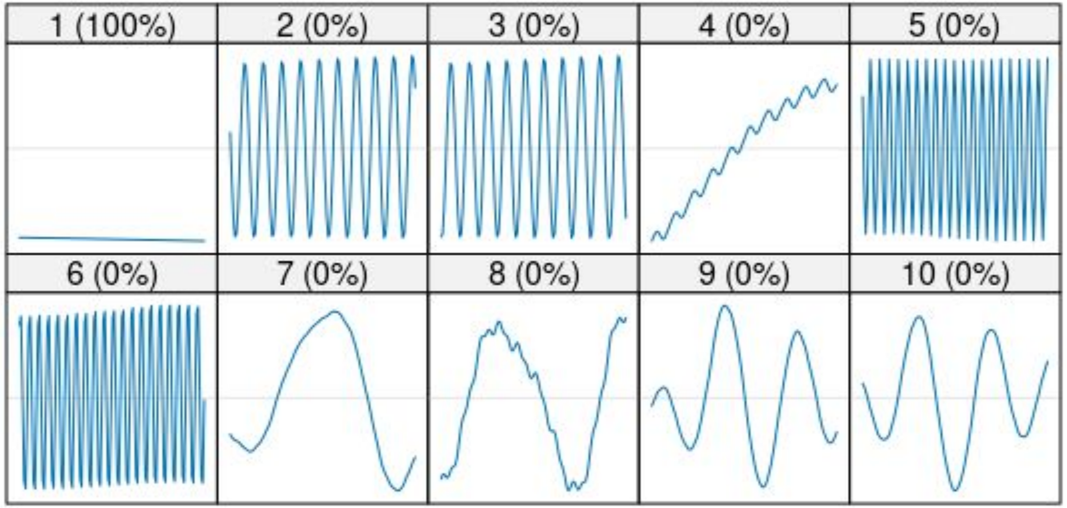
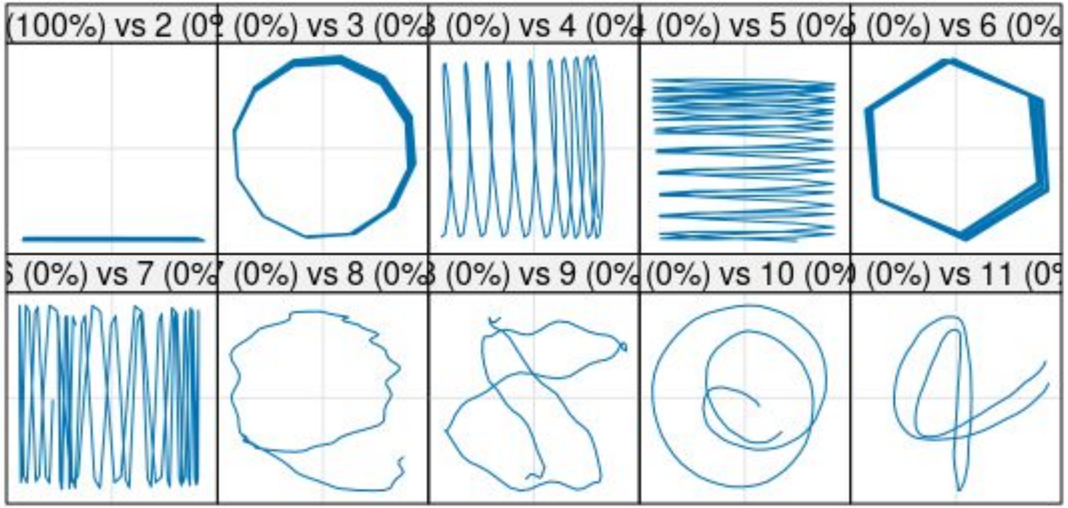
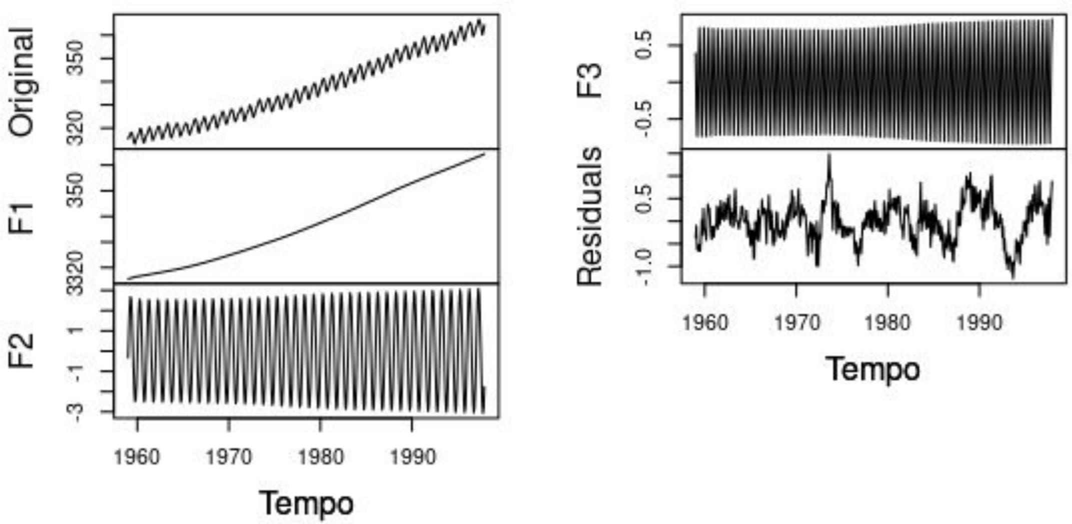


Gráfico dos pares dos dez primeiros autovetores

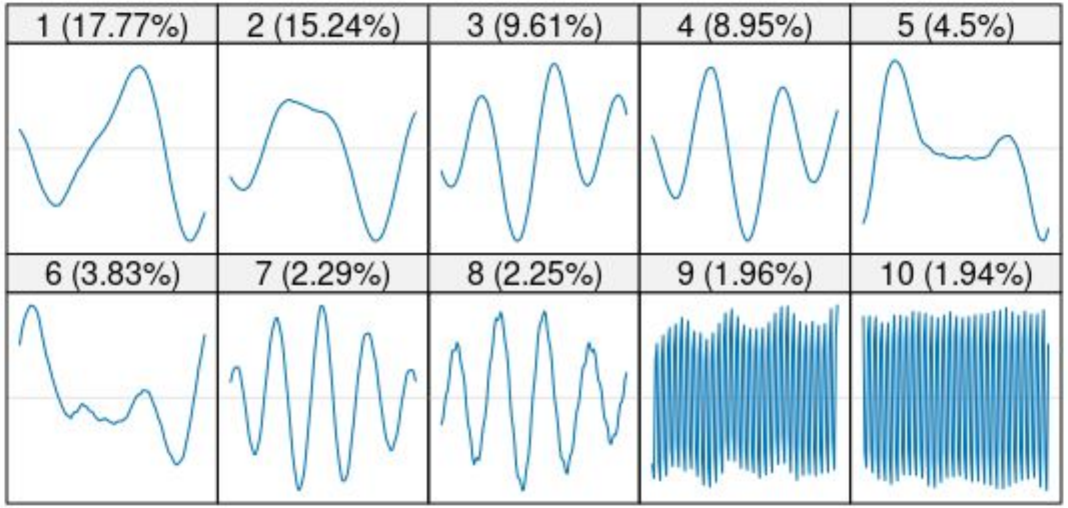


Série CO2 reconstruída

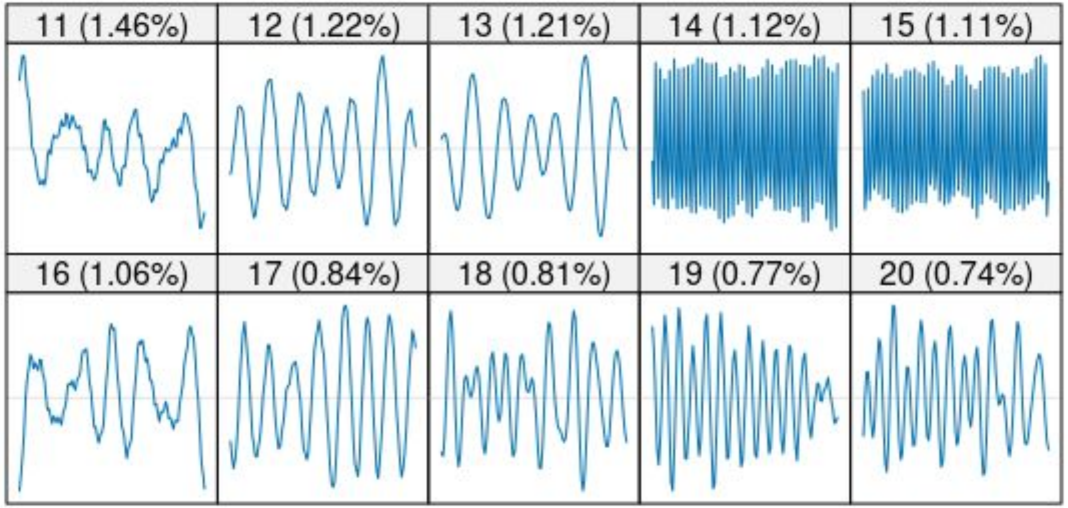


Exercício: encontrar componentes periódicos ocultos nos resíduos da série CO2

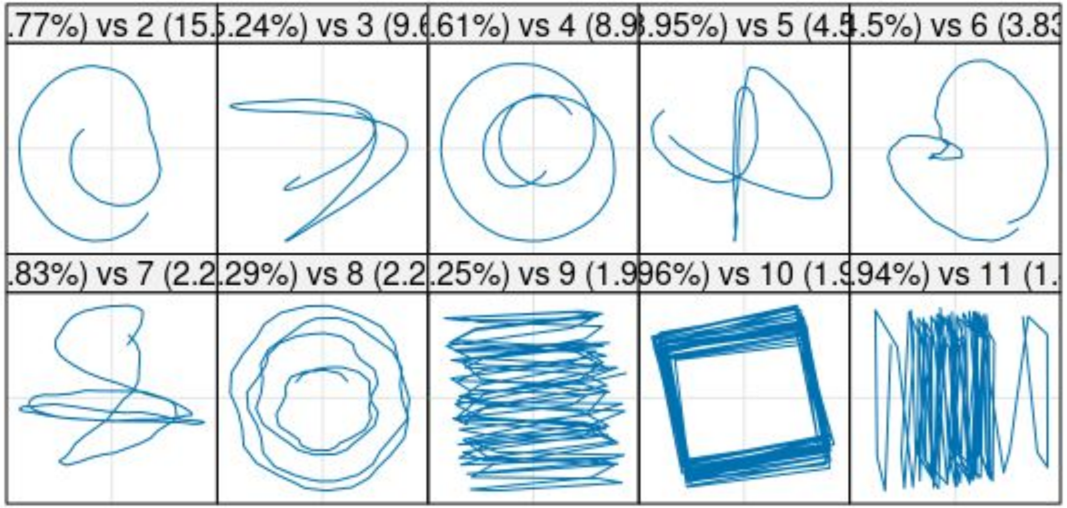
Gráfico dos dez primeiros autovetores da série de resíduos



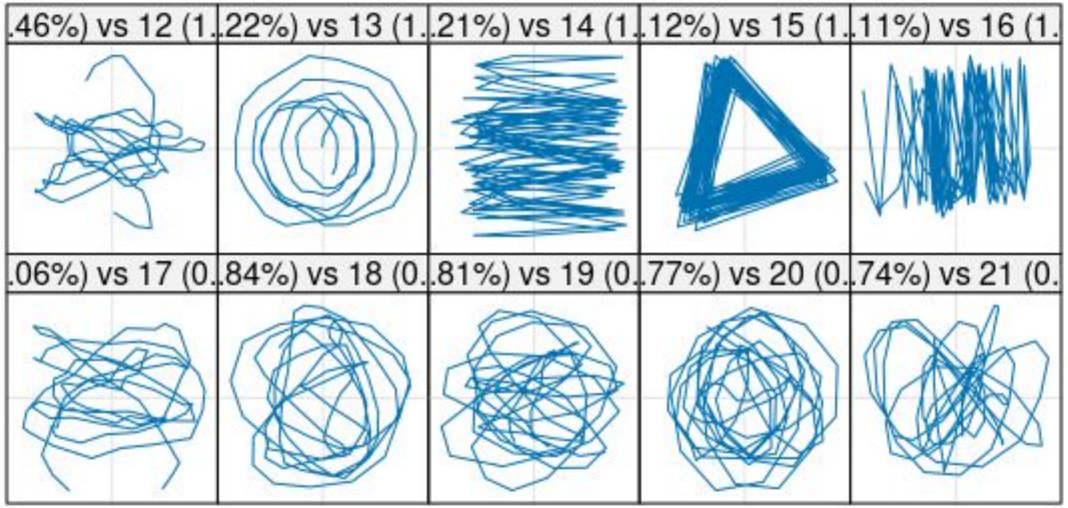
Autovetores do 11 ao 20



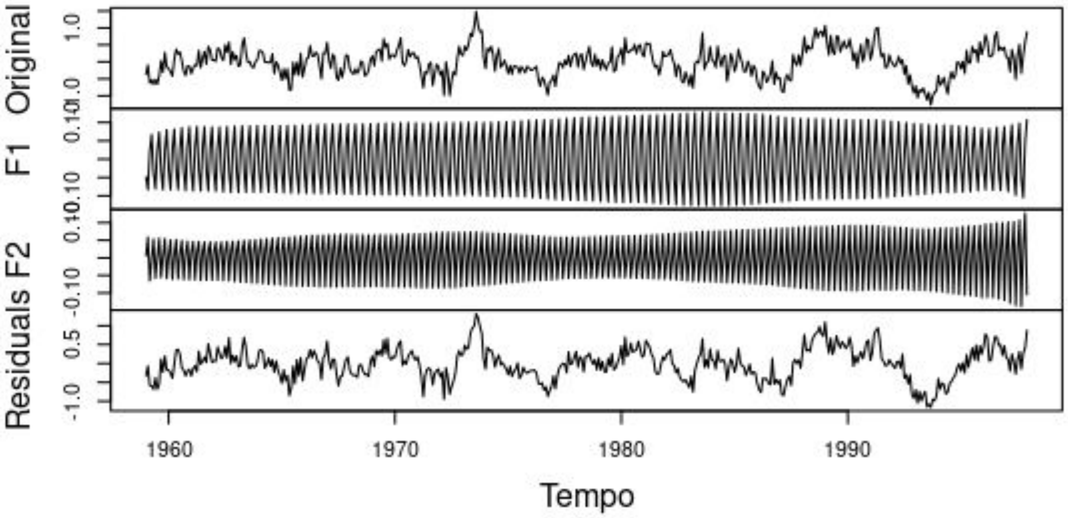
Pares de autovetores de 1 a 10 dos resíduos



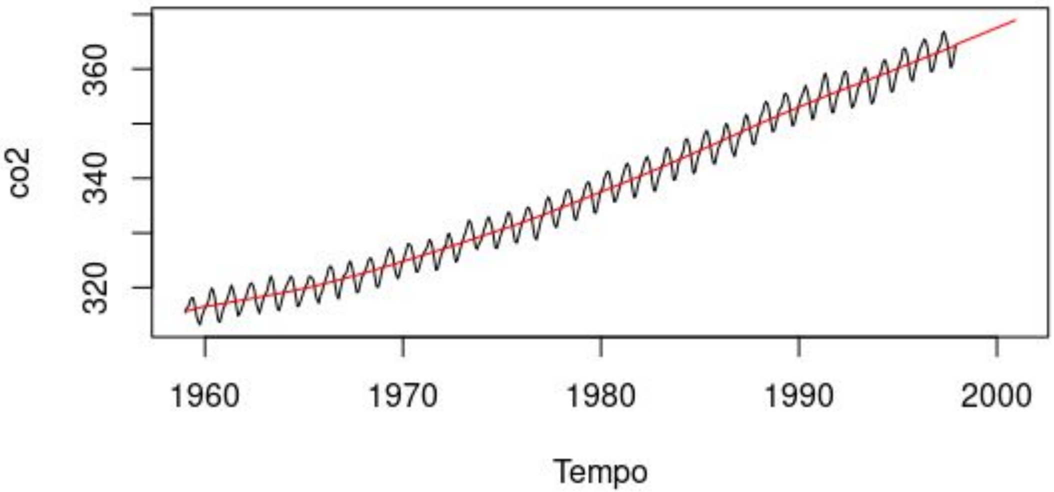
Pares do 11 ao 20



Série de resíduos reconstruída



Previsões baseada no componente de tendência



Obrigado pela atenção!