

DECOMPOSIÇÃO DE SÉRIES TEMPORAIS UTILIZANDO OS ALGORITMOS SSA e SSA RECURSIVO

Guilherme Santos Martins * Mateus Giesbrecht *

* Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, SP, (e-mails: gmartins@dsif.fee.unicamp.br, mateus@fee.unicamp.br).

Abstract: In order to extract temporal characteristics, find the trend and seasonality and eliminate noise from a time series, one of the ways is to use a time series decomposition technique. In this work, two time series decomposition methods are discussed. The first performs the SSA algorithm for time series decomposition using a more usual approach, in which eigenvalues and eigenvectors are calculated in batch from the whole time series. In the second approach, the eigenvalues and eigenvectors are updated recursively. The methods were implemented in MatLab[®] and the computational cost between the algorithms was compared. The results illustrate that the recursive SSA approach has a lower computational cost compared to the batch method.

Resumo: Uma das maneiras de extrair as características temporais, encontrar a tendência, sazonalidade e eliminar o ruído de uma série temporal, é utilizar uma técnica de decomposição de séries temporais. Neste trabalho, são apresentados dois métodos de decomposição de séries temporais. O primeiro realiza o algoritmo SSA para a decomposição da série temporal usando uma abordagem mais usual, em que os autovalores e autovetores são calculados a partir da série temporal em batelada. Na segunda abordagem, os autovalores e autovetores são atualizados recursivamente. Os métodos foram implementados em MatLab[®] e os custos computacionais dos algoritmos foram comparados. Os resultados ilustram que a abordagem SSA recursiva obteve custo computacional menor e com reconstrução da série temporal tão boa quanto à do método em batelada.

Keywords: SSA; Recursive SSA; Time series; Time Series Decomposition

Palavras-chaves: SSA; SSA Recursivo; Séries Temporais; Decomposição Séries Temporais

1. INTRODUÇÃO

Séries temporais são um conjunto de observações feitas em sequência ao longo do tempo, podendo ser monovariáveis ou multivariáveis (Giesbrecht e Bottura, 2011), (Chen et al., 2021). Neste contexto, a predição de séries temporais é uma técnica muito utilizada para prever os dados futuros com base no conhecimento do passado (Rodrigues Júnior e de Oliveira Serra, 2021). Predições de séries temporais apresentam vasta gama de aplicações e estão ligadas à engenharia, finanças, meteorologia, energia, reconhecimento de padrões dentre outros.

Atualmente, modelos estatísticos, modelos de inteligência artificial(AI) e modelos híbridos são os mais utilizados para o problema de predição de séries temporais. Embora vários trabalhos tenham abordado a predição de séries temporais utilizando modelos estatísticos e de AI, no cenário atual estão surgindo novas técnicas que são combinadas com esses modelos, a fim de melhorar a predição futura. Uma dessas técnicas é a decomposição de séries temporais. A partir

da decomposição de séries temporais é possível extrair características importantes, tais como tendência, sazonalidade e ruído, o que pode resultar em melhor desempenho na predição (Rodrigues Júnior e de Oliveira Serra, 2021), (Cassiano, 2014).

As principais técnicas de decomposição de séries temporais encontradas na literatura são a Transformação Wavelet (WT), Análise Componentes Principais (PCA), Decomposição em Modo Empírico (EMD), Análise Singular Espectral (SSA) dentre outras (Das et al., 2022).

1.1 Objetivos do trabalho

O objetivo deste trabalho é realizar a decomposição e a reconstrução de séries temporais utilizando a técnica SSA tradicionalmente encontrada na literatura (Golyandina e Zhigljavsky, 2013). Neste artigo também é utilizada uma variante recursiva do SSA, incorporando os dados de medição mais recentes no conjunto de observações (Rodrigues Júnior e de Oliveira Serra, 2021). Por fim, é realizada uma comparação em termos de custo computacional entre os dois métodos, sendo essa a contribuição desse artigo.

* O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Este artigo está organizado da seguinte maneira: na Seção 2 é apresentada a base de dados utilizada, na Seção 3 é apresentado o método SSA tradicional, na Seção 4 é apresentado o método SSA recursivo, na Seção 5 é apresentada a metodologia, na Seção 6 os resultados e na Seção 7 é apresentada a conclusão.

2. BASE DE DADOS

A base de dados utilizada no desenvolvimento deste trabalho está disponível em (UCI, 2008). Os dados são referentes à medição da qualidade do ar em uma cidade da Itália, mas para este trabalho foram coletados somente os dados da série temporal de temperatura em grau Celsius. A medição de temperatura consiste em um conjunto de 720 dados entre as datas de 1/7/2004 às 00h até 31/07/2004 às 00h, medidos no intervalo de 1 hora. Mais detalhes podem ser vistos em (UCI, 2008).

3. MÉTODO SSA TRADICIONAL

A técnica SSA é um método que vem sendo utilizado em muitas áreas científicas, tais como na análise de séries temporais, economia, matemática financeira, previsões, dentre outros (Bojang et al., 2020) (Cassiano, 2014). Essa técnica decompõe a série temporal em componentes que podem ser identificados como tendência, oscilações e ruído. Um dos grandes benefícios em se utilizar o SSA é que a técnica é considerada não paramétrica, ou seja, não é necessário realizar uma análise a priori do modelo (Cassiano, 2014). Outra vantagem dessa técnica é poder ser utilizada para análise em séries temporais longas ou curtas (Rodrigues et al., 2018).

As quatro fases do SSA são:

- Incorporação (do inglês Embedding): essa etapa consiste em decompor a série temporal em uma matriz de Hankel, conhecida também como matriz trajetória;
- Decomposição em Valores Singulares (do inglês Singular Value Decomposition ou SVD): nessa etapa é utilizada a SVD para decompor a matriz trajetória;
- Agrupamento: essa etapa envolve agrupar a série temporal e identificar as componentes de tendência, sazonal e ruído, com o objetivo de eliminar o ruído;
- Média Diagonal: essa etapa consiste em reconstruir as subséries obtidas a partir da média diagonal.

As etapas são detalhadas a seguir.

1º) Incorporação:

Considere uma série temporal P -dimensional dada por (x_1, \dots, x_N) , em que N é o número de observações. Além disso, considere o tamanho da janela, L , sendo um número inteiro no intervalo de $1 < L < N$, sendo que geralmente $L \leq N/2$. Por fim, $K = N - L + 1$ é um valor inteiro que representa o número de colunas da matriz de trajetória X . Deste modo, é possível formar a matriz trajetória a partir da série temporal original, conforme é expresso na equação (1):

$$X = [X_1 : \dots : X_K] = (x_{ij})_{ij}^{L,K} = \begin{bmatrix} x_1 & x_2 & x_3 & \dots & x_K \\ x_2 & x_3 & x_4 & \dots & x_{K+1} \\ x_3 & x_4 & x_5 & \dots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \dots & x_N \end{bmatrix}, \quad (1)$$

em que X_i é a i -ésima coluna da matriz X , que contém os componentes da série do instante i ao instante $i + L - 1$. Observa-se na equação (1) que a matriz trajetória X é uma matriz blocos de de Hankel.

2º) Decomposição em valores singulares (SVD):

Nessa etapa é realizada a SVD da matriz de trajetória X . Seja:

$$S = XX^T. \quad (2)$$

Os autovalores de S , que são o quadrado dos valores singulares de X , são tomados em ordem decrescente de magnitude, sendo representados como $\lambda_1 \geq \dots \geq \lambda_L \geq 0$, enquanto que $U_1 \geq \dots \geq U_L \geq 0$ são os autovalores de S , correspondente a esses autovalores. Esses também são os vetores ortonormais à esquerda de X . Os vetores ortonormais à direita de X , denotados por V_k , podem ser calculados conforme a equação (3):

$$V_k = \frac{X^T U_k}{\sqrt{\lambda_k}}. \quad (3)$$

Considerando que d seja o número de autovalores diferentes de zero, partindo do princípio que X_k seja conforme a equação (4):

$$X_k = \sqrt{\lambda_k} U_k V_k^T, \quad (4)$$

em que $1 \leq k \leq d$, a matriz X pode ser expressa como:

$$X = X_1 + \dots + X_d. \quad (5)$$

A coleção $(\sqrt{\lambda_k}, U_k, V_k)$ é denominada como a k -ésima autotripla da SVD.

3º) Agrupamento:

A partir da equação (5), o procedimento de agrupamento particiona o conjunto de índices $\{1, \dots, d\}$ em m subconjuntos disjuntos, $\{I_1, \dots, I_m\}$, utilizando método de agrupamento, deste modo $m \leq d$. Alguns desses métodos foram explorados em (Cassiano, 2014).

As matrizes resultantes são calculadas para os grupos $I = I_1, \dots, I_m$ e a equação (5) resulta na seguinte decomposição:

$$X = X_{I_1} + \dots + X_{I_m}. \quad (6)$$

Observa-se que a partir da etapa de agrupamento é possível diminuir o número de componentes da matriz X .

4º) Média diagonal:

A média diagonal é capaz de reconstruir a série temporal original a partir das subséries obtidas no passo anterior. A média diagonal é expressa na equação (7).

$$y_\nu = \begin{cases} \frac{1}{k} \sum_{i=1}^{\nu} y_{i,\nu-i+1} & \text{se } 1 \leq \nu \leq L^* \\ \frac{1}{L^*} \sum_{i=1}^{L^*} y_{i,\nu-i+1} & \text{se } L^* \leq \nu \leq K^* \\ \frac{1}{N-\nu+1} \sum_{i=\nu-K^*+1}^{N-K^*+1} y_{i,\nu-i+1} & \text{se } K^* < \nu \leq N \end{cases} \quad (7)$$

em que $L^* = \min(L, K)$ e $K^* = \max(L, K)$.

Uma das desvantagens da técnica SSA é o custo computacional do SVD da matriz de trajetória, principalmente quando se tem interesse em atualizar os valores dos componentes com a chegada de novas observações. Para resolver esse problema, pode ser usada uma nova abordagem, em que o método SSA é realizado recursivamente, a fim de melhorar o custo computacional desse algoritmo. O método é detalhado na próxima seção.

4. MÉTODO SSA RECURSIVO

Nesta seção é apresentado o desenvolvimento do método SSA recursivo (SSA-R) tomando como referência o procedimento realizado em (Rodrigues Júnior e de Oliveira Serra, 2021), que por sua vez foi baseado nas referências (Erdogmus et al., 2004) e (Mirmomeni et al., 2011). O objetivo principal é atualizar os autovalores e autovetores da matriz S_k recursivamente, sem a necessidade de recorrer ao algoritmo de SVD a cada vez que um novo dado se torna conhecido.

A princípio os autores em (Erdogmus et al., 2004) propõem um método PCA recursivo usando a técnica de perturbação das matrizes de autovalores e autovetores. Observando a similaridade entre esse método e o método SSA, foi feita uma adaptação para que o método SSA tenha característica recursiva. O desenvolvimento do método SSA recursivo é mostrado a seguir.

Considere que no instante k foi observado um novo vetor de dados P -dimensional e que $X_k \in \mathbb{R}^{L \times P}$ seja a última coluna de uma matriz de trajetória similar à apresentada na equação (1), contendo em sua última posição o novo dado observado. A partir da equação (2) nota-se que, no instante k , a matriz S_k é dada por:

$$S_k = \sum_{i=1}^k X_i X_i^T = S_{k-1} + X_k X_k^T, \quad (8)$$

A partir das propriedades de álgebra linear, se Q_k e Λ_k são a matriz ortogonal de autovetores e a matriz diagonal de autovalores, respectivamente, da matriz S_k , é possível obter as seguintes expressões:

$$S_k = Q_k \Lambda_k Q_k^T, \quad (9)$$

$$S_{k-1} = Q_{k-1} \Lambda_{k-1} Q_{k-1}^T, \quad (10)$$

deste modo, substituindo as equações (9) e (10) na equação (8), obtém-se:

$$Q_k \Lambda_k Q_k^T = X_k X_k^T + Q_{k-1} \Lambda_{k-1} Q_{k-1}^T. \quad (11)$$

Definindo $\alpha_k \triangleq Q_{k-1}^T X_k$, pela ortogonalidade de Q_{k-1} tem-se que $X_k = Q_{k-1} \alpha_k$ e a equação (11) pode ser expressa como:

$$Q_k \Lambda_k Q_k^T = Q_{k-1} [\Lambda_{k-1} + \alpha_k \alpha_k^T] Q_{k-1}^T. \quad (12)$$

Sejam V_k e D_k as matrizes de autovetores e de autovalores de $\Lambda_{k-1} + \alpha_k \alpha_k^T$. Com isso, é possível definir a seguinte expressão:

$$\Lambda_{k-1} + \alpha_k \alpha_k^T = V_k D_k V_k^T, \quad (13)$$

substituindo a equação (13) na equação (12), obtém-se:

$$Q_k \Lambda_k Q_k^T = Q_{k-1} V_k D_k V_k^T Q_{k-1}^T. \quad (14)$$

Comparando os lados da equação (14), é possível obter a atualização recursiva das matrizes de autovalores e autovetores, que são expressas nas equações (15) e (16), respectivamente.

$$Q_k = Q_{k-1} V_k, \quad (15)$$

$$\Lambda_k = D_k. \quad (16)$$

Porém, ainda não é possível encontrar a atualização das matrizes V_k e D_k , expressas nas equações (15) e (16). Para que se torne possível, é necessário encontrar a decomposição de $\Lambda_{k-1} + \alpha_k \alpha_k^T$, mencionada anteriormente. Uma forma de determinar essa decomposição é fazer o uso da análise da matriz de perturbação de 1^a ordem, desenvolvida a seguir.

Análise de perturbação para determinar V_k e D_k

Ao se analisar a equação (13) é possível dizer que, para um k suficientemente grande, o termo $\Lambda_{k-1} + \alpha_k \alpha_k^T$ é próximo de D_k e que, portanto, V_k é próximo da matriz identidade I . A matriz $\alpha_k \alpha_k^T$ é dita matriz diagonal de perturbação de Λ_{k-1} , pois é pequena em relação ao outro termo. Deste modo, é possível expressar D_k e V_k , conforme as equações (17) e (18), respectivamente:

$$D_k = \Lambda_{k-1} + P_\Lambda, \quad (17)$$

$$V_k = I + P_V, \quad (18)$$

em que P_Λ e P_V são pequenas perturbações. Deste modo, ao encontrar as matrizes de perturbações, o problema da autodecomposição de $\Lambda_{k-1} + \alpha_k \alpha_k^T$ (isto é, encontrar as matrizes de D_k e V_k) será resolvido.

Considere $\Lambda = \Lambda_{k-1}$ e substituindo as equações (17) e (18) no termo $V_k D_k V_k^T$, obtém-se:

$$V_k D_k V_k^T = (I + P_V)(\Lambda + P_\Lambda)(I + P_V)^T, \quad (19)$$

$$V_k D_k V_k^T = \Lambda + \Lambda P_V^T + P_\Lambda + P_\Lambda P_V^T + P_V \Lambda + P_V \Lambda P_V^T + P_V P_\Lambda + P_V P_\Lambda P_V^T \quad (20)$$

$$= \Lambda + P_\Lambda + D_k P_V^T + P_V D_k + P_V \Lambda P_V^T + P_V P_\Lambda P_V^T \quad (21)$$

Considerando que os termos $P_V \Lambda_{k-1} P_V^T + P_V P_\Lambda P_V^T$ sejam desprezíveis por serem formas quadráticas da matriz P_V , que por sua vez é pequena, e substituindo a equação (21) na equação (13), obtém-se:

$$\Lambda_{k-1} + \alpha_k \alpha_k^T \approx \Lambda_{k-1} + P_\Lambda + D_k P_V^T + P_V D_k, \quad (22)$$

simplificando a equação (22), tem-se:

$$\alpha_k \alpha_k^T \approx P_\Lambda + D_k P_V^T + P_V D_k. \quad (23)$$

Como a matriz V_k é ortogonal, $V_k V_k^T = I$. Deste modo, substituindo a equação (18) em $V_k V_k^T = I$ e considerando que $P_V P_V^T \approx 0$, é obtida a seguinte expressão:

$$P_V = -P_V^T. \quad (24)$$

Sabendo-se que P_V é anti-simétrico, P_Λ e D_k são matrizes diagonais, é possível obter soluções para as matrizes de perturbações P_V e P_Λ , analisando a equação (23), conforme é expresso na equação (25).

$$\begin{cases} \alpha_i^2, & = (i, i)_{-esimo} \text{ elemento de } P_\Lambda \\ \frac{\alpha_i \alpha_j}{\lambda_j + \alpha_j^2 - \lambda_i - \alpha_i^2} & = (i, j)_{-esimo} \text{ elemento de } P_V, i \neq j \\ 0, & = (i, i)_{-esimo} \text{ elemento de } P_V \end{cases} \quad (25)$$

em que λ_i, λ_j são os elementos da matriz autovalor de Λ_{k-1} . Com isso, a partir de α_k , que pode ser completamente calculado com informações sobre a decomposição no instante $k-1$ e a nova observação, é possível determinar completamente as matrizes de perturbação e consequentemente os autovalores e autovetores da matriz S_k . Deste ponto em diante o algoritmo SSA segue como no caso em batelada. Mais detalhes podem ser vistos nas referências (Rodrigues Júnior e de Oliveira Serra, 2021), (Erdogmus et al., 2004) e (Mirmomeni et al., 2011).

O Algoritmo 1 resume o procedimento realizado para o método SSA-R.

Algoritmo 1 SSA-R.

- 1: 1) Inicialize o sistema. Para isso:
 - 2: a) Insira a série temporal
 - 3: b) escolha o tamanho da janela de inicialização (L)
 - 4: c) escolha o número de amostras de inicialização
 - 5: d) obtenha a matriz de trajetória X
 - 6: 2) Para cada novo dado de entrada:
 - 7: e) Obtenha o vetor de entrada X_k
 - 8: f) Calcule $\alpha_k = Q_{k-1} X_k$
 - 9: g) A partir da equação (25), Determine P_V e P_Λ
 - 10: h) A partir da equação (17), calcule D_k
 - 11: i) A partir da equação (18), calcule V_k
 - 12: j) A partir da equação (15), atualize a matriz de autovetores
 - 13: l) A partir da equação (16), atualize a matriz diagonal de autovalores
 - 14: m) Proceda os passos do método SSA descritos entre as equações (3) e (7)
 - 15: n) Retorne para o passo 7.
-

5. METODOLOGIA

A implementação da técnica SSA em batelada foi executada para decompor a série temporal conforme desenvolvido na Seção 3. Vale destacar que novos dados de

medição mais recentes foram incorporados no conjunto de observações a cada iteração, repetindo a SVD a cada novo dado. Foi utilizado um tamanho de janela, L, fixo, igual a três.

Para fins experimentais foi usada a base de dados apresentada na Seção 2 e não foi utilizada nenhuma técnica de pré-processamento. Foram feitos dois experimentos. O primeiro, consiste em uma série temporal com as 100 primeiras medições de temperatura, em que os dados de inicialização foram variados de 10 a 40 com um intervalo de 10. No segundo experimento foi usada uma série temporal de 720 medições, em que os dados de inicialização foram variados de 20 a 120 com um intervalo de 50.

Cada experimento foi executado 1000 vezes para se calcular a média do tempo de execução por iteração.

A partir do desenvolvimento da técnica SSA recursiva apresentada na Seção 4, foram feitos os mesmos experimentos da técnica SSA tradicional a fim de comparar os desempenhos de ambos os métodos.

Todos os cálculos foram programados em MatLab[®] e executados utilizando um computador com as seguintes configurações: Workstation Monoprocessada Xeon E3-1270V6 HT, Modelo Deer Flat Intel Xeon E3-1270V6 (4-core, 3.8 GHz, 8 MB), RAM 64 GB, SSD 240 GB, HD 1 TB.

6. RESULTADOS

Na Tabela 1 é mostrado o tempo médio de execução por iteração usando a série temporal de temperatura. Para esse experimento, foram usadas as 100 primeiras medições. Já na Tabela 2 é mostrado o tempo médio de execução por iteração usando a série temporal de temperatura usando um conjunto de 720 pontos.

Na Figura 1 é ilustrado o tempo médio de execução por iteração a medida em que os dados são incorporados para serem usados na decomposição para o método SSA em batelada, enquanto que na Figura 2 é ilustrado o tempo médio de execução por iteração a medida em que os dados são incorporados para serem usados na decomposição para o método SSA recursivo. Para os experimentos foram usados 100 medições.

Nas Tabelas 3 e 4 são mostrados os MSEs referentes a cada número de dados de inicialização para um conjunto de 100 pontos e outro de 720 pontos, respectivamente, usando os métodos SSA recursivo e SSA em batelada.

Tabela 1. Tempo médio de execução por iteração usando um conjunto de 100 medições.

Método	Número de amostras de inicialização	Tempo médio por iteração (s)	Desvio Padrão
SSA Batelada	10	8.62×10^{-5}	3.18×10^{-5}
SSA Recursiva	10	1.41×10^{-5}	1.82×10^{-5}
SSA Batelada	20	8.02×10^{-5}	2.80×10^{-5}
SSA Recursiva	20	1.27×10^{-5}	1.76×10^{-5}
SSA Batelada	30	7.49×10^{-5}	2.45×10^{-5}
SSA Recursiva	30	1.12×10^{-5}	1.35×10^{-6}
SSA Batelada	40	6.82×10^{-5}	2.10×10^{-5}
SSA Recursiva	40	9.68×10^{-6}	1.55×10^{-6}

Tabela 2. Tempo médio de execução por iteração usando um conjunto de 720 medições.

Método	Número de amostras de inicialização	Tempo médio por iteração (s)	Desvio Padrão
SSA Batelada	20	4.64×10^{-4}	2.44×10^{-4}
SSA Recursiva	20	2.90×10^{-5}	5.96×10^{-6}
SSA Batelada	70	4.54×10^{-4}	2.24×10^{-4}
SSA Recursiva	70	3.52×10^{-5}	6.32×10^{-6}
SSA Batelada	120	4.43×10^{-4}	2.06×10^{-4}
SSA Recursiva	120	3.19×10^{-5}	5.49×10^{-6}

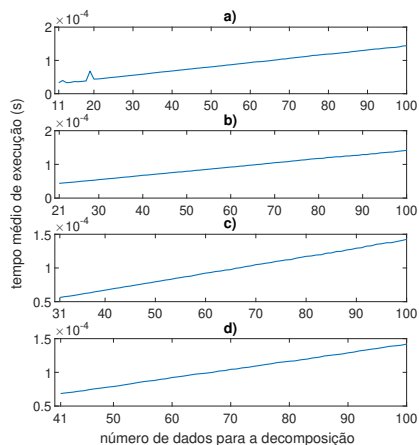


Figura 1. Tempo médio de execução por iteração a medida que os dados são incorporados para serem usados na decomposição utilizando o método SSA em batelada com 10 dados de inicialização (a), com 20 dados de inicialização (b), com 30 dados de inicialização (c) e com 40 dados de inicialização

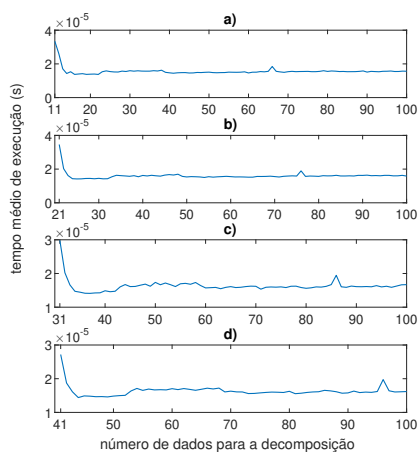


Figura 2. Tempo médio de execução por iteração a medida que os dados são incorporados para serem usados na decomposição utilizando o método SSA recursivo com 10 dados de inicialização (a), com 20 dados de inicialização (b), com 30 dados de inicialização (c) e com 40 dados de inicialização

Na Figura 3 é ilustrada a série temporal original e a série temporal reconstruída com 10 e 30 dados de inicialização, utilizando o conjunto de 100 medições.

Tabela 3. MSE referente a cada número de dados de inicialização usando um conjunto de 100 medições.

Método	Inicialização	MSE
SSA Batelada	10	2.6632×10^{-29}
SSA Recursiva	10	0.5589
SSA Batelada	20	2.6632×10^{-29}
SSA Recursiva	20	6.7977×10^{-4}
SSA Batelada	30	2.6632×10^{-29}
SSA Recursiva	30	2.5874×10^{-4}
SSA Batelada	40	2.6632×10^{-29}
SSA Recursiva	40	2.9220×10^{-5}

Tabela 4. MSE referente a cada número de dados de inicialização usando um conjunto de 720 medições.

Método	Inicialização	MSE
SSA Batelada	20	3.2764×10^{-29}
SSA Recursiva	20	0.3581
SSA Batelada	70	3.2764×10^{-29}
SSA Recursiva	70	3.9285×10^{-4}
SSA Batelada	120	3.2764×10^{-29}
SSA Recursiva	120	2.63×10^{-5}

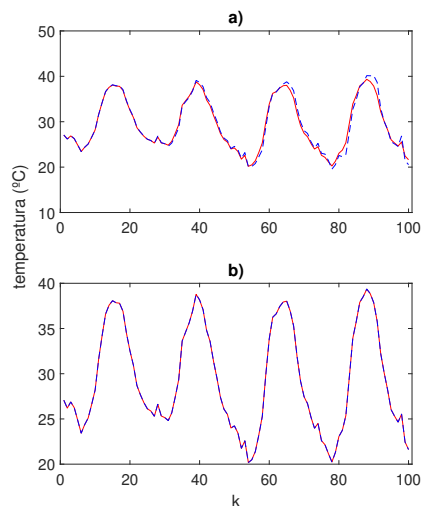


Figura 3. Série temporal original (linha vermelha) e reconstruída (linha tracejada azul) para a temperatura utilizando o método SSA recursivo, com 10 dados de inicialização (a) e com 30 dados de inicialização (b) em um conjunto de 100 medições

6.1 Discussão

Nas Tabelas 1 e 2 é possível notar que a abordagem recursiva SSA obteve um custo computacional menor se comparado ao método SSA em batelada. Esse comportamento é esperado uma vez que os cálculos envolvendo o método SSA recursivo são menos complexos do que o método SSA padrão, pois para o método SSA em batelada é necessário calcular todas as etapas, inclusive a SVD, o que torna o custo computacional maior. Em outras palavras, ao se analisar uma série temporal em que há muitos dados o método SSA recursivo leva grande vantagem se comparado ao do método SSA padrão.

Nota-se na Figura 1 que o tempo médio de execução por iteração aumenta no método em batelada a medida em

que novos dados são incorporados para a decomposição. Isso ocorre pois a matriz trajetória X aumenta conforme novos dados são incorporados, aumentando o custo computacional do cálculo da matriz S . Na Figura 2 nota-se que o tempo de execução por iteração do método recursivo praticamente é constante a medida em que novos dados são incorporados. Isso ocorre pois os cálculos envolvendo o método SSA recursivo não se alteram em função do número de amostras.

Nas Tabelas 3 e 4 é possível notar que o MSE do método recursivo diminui a medida que a quantidade de dados de inicialização aumenta, conforme foi mencionado na Seção 4 em que discutiu-se que é necessária uma quantidade de dados de inicialização suficiente para que seja possível reconstruir a série temporal original. Nota-se na Figura 3, que a série temporal reconstruída utilizando 10 dados de inicialização não obteve resultados tão bons quanto quando foram utilizados 30 dados de inicialização. Portanto, o projetista deve se atentar a esse detalhe e escolher o número de dados de inicialização suficiente para que consiga reconstruir a série temporal original. Além disso, ao se analisar os MSEs dos métodos SSA recursivo e SSA em batelada nas Tabelas 3 e 4 é possível notar que o MSE do método SSA em batelada é menor se comparado ao do método SSA recursivo, porém quando é utilizada uma certa quantidade de dados de inicialização para o método SSA recursivo, o MSE obtido é muito baixo sendo na ordem de 10^{-4} , o que pode ser considerado, dependendo da aplicação, aproximadamente zero. Deste modo, o projetista deve se atentar a esse detalhe, escolhendo uma certa quantidade de dados de inicialização para que o MSE seja baixo e combinando a vantagem, do fato, do menor custo computacional do método SSA recursivo se comparado ao do método SSA em batelada.

7. CONCLUSÃO

Neste trabalho, duas metodologias para decomposição de séries temporais foram apresentadas. A primeira metodologia foi a técnica SSA em batelada para a decomposição de séries temporais, usualmente encontrada na literatura, enquanto a segunda técnica é recursiva.

A técnica SSA recursiva apresentou tempo de execução menor comparado com a técnica SSA em batelada e com valores de reconstrução muito próximos ao da série temporal original, desde que um número suficientemente grande de dados de inicialização seja utilizado. Além disso, percebe-se que a medida que o número de dados de inicialização aumenta, o MSE diminui. Portanto, conclui-se que é necessário inicializar o método com dados suficientes para que possa ser realizada uma reconstrução com baixos valores de erro.

Por fim, a partir dos resultados obtidos, foi observado que são necessários no mínimo 30 dados de inicialização usando um conjunto de 100 medições, para que o método SSA recursivo seja capaz de reconstruir a série temporal original com um erro razoável, enquanto que usando um conjunto de 720 medições, foram necessários 120 dados.

Nos próximos passos desta pesquisa, serão explorados algoritmos estatísticos e de aprendizado de máquina (Martins e Giesbrecht, 2020) para a implementação de predição de

séries temporais utilizando a técnica SSA. Nessa aplicação as abordagens em batelada e recursiva serão comparadas, para observar o tempo de execução e o MSE em horizontes de curto e longo prazo.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

REFERÊNCIAS

- Bojang, P.O., Yang, T.C., Pham, Q.B., e Yu, P.S. (2020). Linking singular spectrum analysis and machine learning for monthly rainfall forecasting. *Applied Sciences*, 10(9), 3224.
- Cassiano, K.M. (2014). Análise de séries temporais usando análise espectral singular (ssa) e clusterização de suas componentes baseada em densidade. *Pontifícia Universidade Católica do Rio de Janeiro*.
- Chen, W., Xu, H., Chen, Z., e Jiang, M. (2021). A novel method for time series prediction based on error decomposition and nonlinear combination of forecasters. *Neurocomputing*, 426, 85–103.
- Das, S., Prusty, B.R., Bingsi, K., Panda, K.P., e Panda, G. (2022). Time series decomposition techniques for renewable generation applications. *Sustainable Energy and Technological Advancements*, 847–856.
- Erdogmus, D., Rao, Y.N., Peddaneni, H., Hegde, A., e Principe, J.C. (2004). Recursive principal components analysis using eigenvector matrix perturbation. *EURASIP Journal on Advances in Signal Processing*, 2004(13), 1–8.
- Giesbrecht, M. e Bottura, C.P. (2011). Uma proposta imuno inspirada para a modelagem de séries temporais discretas no espaço estado. *SBAI - Simpósio Brasileiro de Automação Inteligente*.
- Golyandina, N. e Zhigljavsky, A. (2013). *Singular spectrum analysis for time series*. Springer.
- Martins, G.S. e Giesbrecht, M. (2020). Estimação de índice de claridade utilizando método de aoki e observador ótimo de estado com inicialização exata. *Anais da Sociedade Brasileira de Automática*, 2(1).
- Mirmomeni, M., Lucas, C., Araabi, B., Moshiri, B., e Bidar, M. (2011). Recursive spectral analysis of natural time series based on eigenvector matrix perturbation for online applications. *IET signal processing*, 5(6), 515–526.
- Rodrigues, P.C., Lourenço, V., e Mahmoudvand, R. (2018). A robust approach to singular spectrum analysis. *Quality And Reliability Engineering International*, 34(7), 1437–1447.
- Rodrigues Júnior, S.E. e de Oliveira Serra, G.L. (2021). An approach for evolving neuro-fuzzy forecasting of time series based on parallel recursive singular spectrum analysis. *Fuzzy Sets and Systems*.
- UCI, M.L.R. (2008). Base de dados. <http://archive.ics.uci.edu/ml/datasets/Air+Quality/> Acesso em: 14 Abr. 2022.