

4.

Análise Espectral Singular (SSA)

A Análise Espectral Singular (*Singular Spectrum Analysis* - SSA) é uma técnica não paramétrica, isto é, não é preciso conhecer o modelo paramétrico que governa a série temporal de interesse; assim a técnica baseia-se exclusivamente nos dados (Beneki et al., 2009). SSA foi desenvolvida para modelar séries temporais não-lineares e/ou não-estacionárias, assim também como séries pequenas e ruidosas (Hassani, 2007). Esta metodologia incorpora elementos da análise clássica de séries temporais, teoria de probabilidade, estatística multivariada, geometria multivariada, sistemas dinâmicos e conceitos de processamento de sinais (Golyandina et al., 2001). Com isto em mente, a técnica SSA será usada no decorrer desta tese com o propósito de prever a velocidade horária do vento, pois sua característica não paramétrica se torna uma vantagem no processo de modelado e previsão em séries temporais.

O objetivo de SSA é decompor a série temporal original na soma de componentes independentes e interpretáveis, as quais representam a tendência, o comportamento oscilatório (componentes periódicas ou quase-periódicas) e ruído sem estrutura. Uma vez determinadas estas componentes, são selecionadas para reconstruir uma aproximação da série sem ruído, em seguida são usadas para fazer previsão.

SSA é relativamente recente, mas seu surgimento gerou-se com as primeiras publicações de Broomhead & King (1986a, 1986b) e Broomhead et al. (1987); enquanto que as mesmas ideias do método estavam sendo desenvolvidas simultaneamente, mas de forma independente, por diversos grupos de pesquisa na Rússia, Reino Unido e Estados Unidos. SSA é uma ferramenta muito útil que pode ser utilizada para resolver os seguintes problemas: 1) encontrar tendências de diferente resolução; 2) alisamento (suavização); 3) extração de componentes de sazonalidade; 4) extração simultânea de ciclos com pequenos e grandes períodos; 5) extração de periodicidades com amplitudes diferentes; 6) extração simultânea de tendências complexas e periodicidades; 7) encontrar estrutura em série temporais curtas; 8) detecção de mudanças estruturais; e 9) a previsão de uma série temporal univariada ou multivariada.

As áreas de aplicação de SSA são diversas: desde matemática e física até economia e matemática financeira (Carvalho et al., 2012; Hassani et al., 2009), desde climatologia, hidrologia e geofísica até ciências sociais e pesquisa de mercado, assim como em engenharia médica (Sanei et al., 2011), entre outras. Qualquer série aparentemente complexa, com uma potencial estrutura poderia fornecer outro exemplo de uma aplicação bem sucedida de SSA (Golyandina et al., 2001).

Segundo Mahmoudvand et al. (2013), uma revisão feita da literatura sobre SSA mostra que há mais de cem artigos sobre as aplicações de SSA nas diferentes áreas e, na maioria delas, a superioridade de SSA comparada com outras técnicas de análise de séries temporais tem sido demonstrada (ver, por exemplo, Hassani et al., 2009; Hassani e Thomakos, 2010 e referências).

O método básico de SSA para *previsão* está composto por três estágios complementares: *decomposição*, *reconstrução* e *previsão* (A Figura 4.1 proporciona uma visão geral da metodologia). No primeiro estágio, decompõe-se a série em diversas componentes (tendência, oscilações e ruído), no segundo estágio, a série livre de ruído seria reconstruída e assim utilizada no terceiro e último estágio para a previsão de novos pontos de dados. A seguir, serão descritos em detalhes os três estágios supracitados.

4.1.

Metodologia SSA

Considere a série temporal de valores reais diferentes de zero $Y_T = (y_1, \dots, y_T)$ de comprimento T (suficiente). O propósito principal de SSA é a **decomposição da série de interesse em vários componentes aditivos, isto é, uma soma de séries**, de modo que cada componente nesta soma possa ser identificada como uma ou várias *componentes de tendência*, *componentes oscilatórias* (talvez com diferentes amplitudes) e *componentes de ruído*. Seguindo isto por uma **reconstrução** da série original.

O objetivo da SSA básica é uma decomposição da série de interesse em várias componentes aditivas que normalmente pode ser interpretado como componentes de tendência (isto é, lisa e variando lentamente em partes da série), várias componentes oscilatórias (talvez com diferentes amplitudes), e componentes "ruído". SSA não se baseia numa função definida a priori, como por exemplo, a abordagem de Fourier (com base em funções seno e cosseno), mas gera um conjunto de componentes diretamente da série temporal em estudo.

A técnica SSA básica compreende dois estágios complementares: decomposição e reconstrução, os quais incluem duas etapas separadas. No primeiro estágio, decompor a série e, no segundo estágio, reconstruir a série original e usar a série reconstruída (que é sem ruído) para a previsão dos novos pontos de dados num terceiro estágio que pode ser denominado estágio de previsão. Abaixo se apresentará uma breve discussão sobre a metodologia da técnica SSA seguindo principalmente Golyandina et al. (2001, cap. 1 e 2).

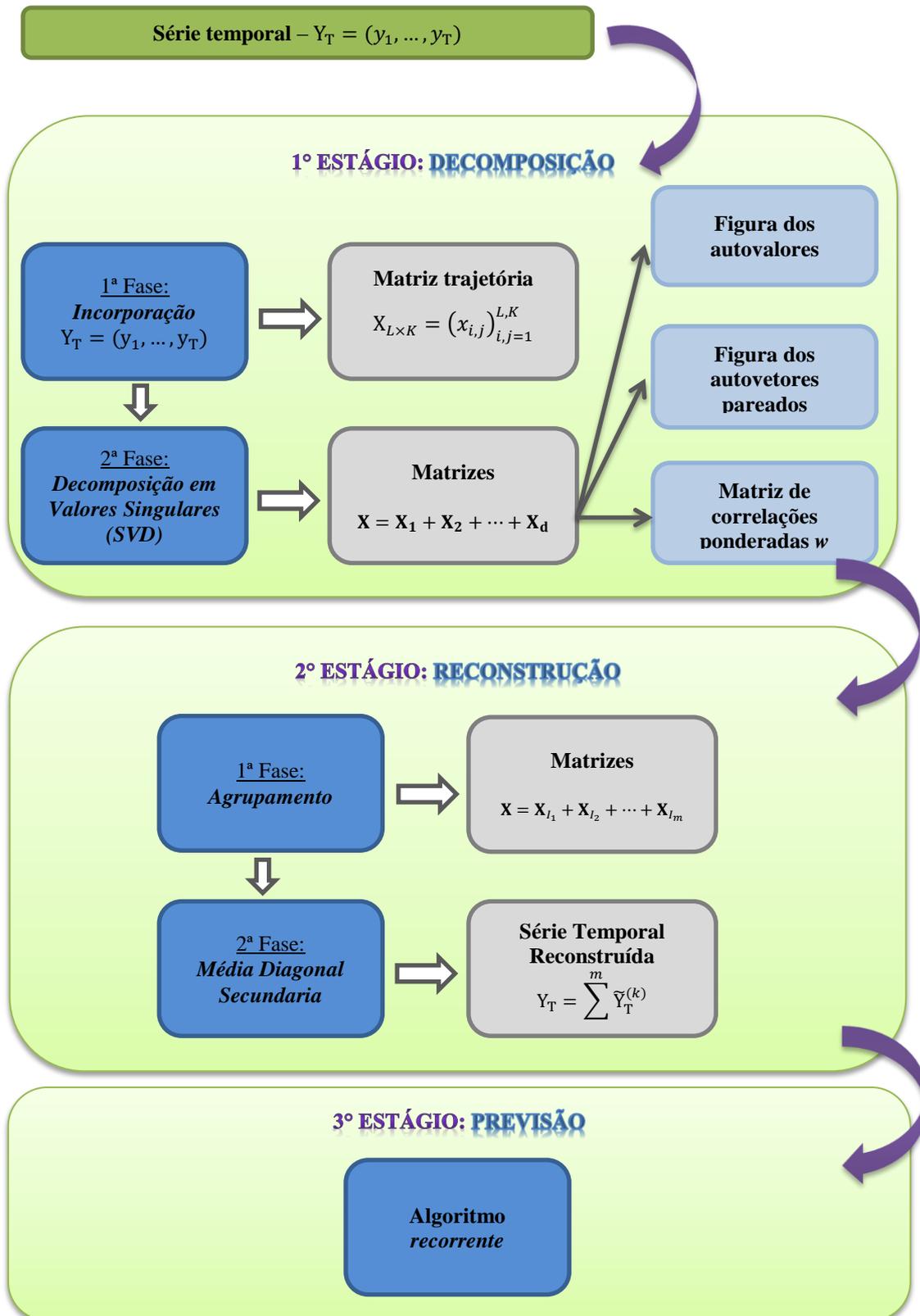


Figura 4.1 – Estágios da Metodologia SSA para previsão.

(Fonte: Elaboração do Autor)

Cabe destacar aqui que, apesar de alguns conceitos probabilísticos e estatísticos serem empregados nos métodos baseados em SSA, não será feita qualquer suposição estatística tal como estacionariedade da série ou normalidade dos resíduos (Golyandina et al., 2001).

4.1.1.

Decomposição

Neste primeiro estágio é executada primeira fase (chamada fase de incorporação), na qual a série unidimensional é representada como uma série multidimensional, cuja dimensão é chamada comprimento da janela. A série temporal multidimensional (que é uma sequência de vetores) forma a trajetória matriz. O único parâmetro (e muito importante) desta fase é o comprimento da janela.

A segunda fase é a decomposição em valores singulares (SVD – *Singular Value Decomposition*) da matriz de trajetória em uma soma de matrizes bi-ortogonais de posto 1, ou seja, a série é subdividida na soma de pequenas componentes interpretáveis. As duas primeiras fases constituem assim o estágio de *decomposição* de SSA básico.

4.1.1.1.

Incorporação

A *incorporação* pode ser considerada como um mapeamento que transfere uma série temporal unidimensional $Y_T = (y_1, y_2, \dots, y_T)$ em uma série multidimensional com dimensão L denominada *window length* ou comprimento da janela. Esta série multidimensional é uma sequência de vetores X_1, \dots, X_K os quais estão constituídos por elementos da série temporal Y_T , em que $X_i = (y_i, \dots, y_{i+L-1})^T \in \mathbb{R}^L$, sendo $K = T - L + 1$ (com $1 \leq i \leq K$), e o qual representa o número de vetores coluna ou subséries defasadas no tempo. O comprimento da janela L é o único parâmetro da fase de *incorporação*, o qual é um número inteiro tal que $2 \leq L \leq T - 1$, suficientemente grande, mas não superior a $T/2$ e inferior a $T/3$, representando a quantidade de componentes em que a série temporal original é decomposta (Golyandina et al., 2001; Hassani, 2007). Não existe uma regra universal para a seleção do comprimento da janela, mas segundo Golyandina & Zhigljavsky,(2013) a decomposição mais detalhada é obtida quando $L \approx T/2$. O resultado deste passo é chamado matriz trajetória:

$$\mathbf{X} = [X_1 : \dots : X_K] = (x_{i,j})_{i,j=1}^{L,K} = \begin{bmatrix} y_1 & y_2 & y_3 & \dots & y_K \\ y_2 & y_3 & y_4 & \dots & y_{K+1} \\ y_3 & y_4 & y_5 & \dots & y_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \dots & y_T \end{bmatrix} \quad (4.1)$$

Pode-se observar que a matriz trajetória \mathbf{X} em (4.1) possui uma estrutura semelhante à de uma matriz de Hankel, o que significa que todos os elementos $\{a_{ij}\}$ ao longo das diagonais paralelas à diagonal secundária são iguais, além de ser uma matriz simétrica. Note-se também que na matriz trajetória a primeira coluna (y_1, \dots, y_L) e a última linha (y_L, \dots, y_T) formam a série original.

Para ilustrar o resultado desta fase se utilizará o exemplo citado em Mahmoudvand et al. (2013). Seja a $Y = [1, 2, 3, 4, 5, 6]$, $T = 6$. Suponha $L = 3$, logo, $K = 6 - 3 + 1 = 4$. A matriz trajetória está dada então por:

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

Observe como dito anteriormente que a primeira coluna $[1, 2, 3]$ e a última linha $[3, 4, 5, 6]$ conformam a série original.

4.1.1.2.

Decomposição em Valores Singulares (SVD)

Nesta fase é feita a *Decomposição em Valores Singulares (SVD – Singular Value Decomposition)* da matriz de trajetória \mathbf{X} em (4.1), a qual é representada como uma soma de matrizes elementares bi-ortogonais de posto 1.

Seja \mathbf{S} o produto entre a matriz trajetória e a sua transposta, $\mathbf{S} = \mathbf{X}\mathbf{X}^T$. Ao aplicar sobre \mathbf{S} a SVD são obtidos os autovalores $\lambda_1, \dots, \lambda_L$ e dispostos em ordem decrescente de acordo com as suas magnitudes ($\lambda_1 \geq \dots \geq \lambda_L \geq 0$), e os autovetores ou vetores singulares U_1, \dots, U_L correspondentes a um sistema de autovetores ortonormais (isto é, $(U_i, U_j) = 0$ para $i \neq j$ (propriedade de ortogonalidade) e $\|U_i\| = 1$ (propriedade de normalização)) associados aos autovalores de \mathbf{S} , conformando U_i assim uma base ortonormal para \mathbb{R}^L . (U_i, U_j) é o produto interno dos vetores U_i e U_j , e U_i é a norma do vetor U_i . Desta forma, a decomposição em autovalores da matriz de trajetória \mathbf{X} pode ser escrita como:

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_L \quad (4.2)$$

sendo $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$ ($i = 1, \dots, d$). Além disso, tem-se que $d = \max(i, \text{tal que } \lambda_i > 0) = \text{posto}(\mathbf{X})$, isto é, d representa o número de autovalores diferentes de zero da matriz \mathbf{S} ($d \leq L$). Ao ser considerada a transformação $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$, ($i = 1, \dots, d$), a qual é a representação das componentes principais da matriz trajetória, a SVD da matriz trajetória \mathbf{X} pode ser escrita como:

$$\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_d = \sum_{i=1}^d \mathbf{X}_i \quad (4.3)$$

\mathbf{X}_i representa uma matriz de posto unitário, normalmente conhecidas como matriz elementar. U_i (na literatura de SSA eles são chamados de "*Fator de Funções Ortogonais Empíricas*" ou simplesmente EOFs) e V_i (frequentemente chamado de "componentes principais") representam os autovetores esquerdo (U_i) e direito (V_i), respectivamente, da matriz de trajetória \mathbf{X} . A coleção $(\sqrt{\lambda_i}, U_i, V_i)$ é chamada a i -ésima autotripla da matriz \mathbf{X} , $\sqrt{\lambda_i}$ ($i = 1, \dots, d$) são os valores singulares da matriz \mathbf{X} e o conjunto $\{\sqrt{\lambda_i}\}$ é chamado o espectro da matriz \mathbf{X} . Se todos os autovalores têm multiplicidade unitária, então a expansão (4.3) é definida de forma única. Caso contrário, se houver pelo menos um autovetor com multiplicidade maior do que 1, então existe liberdade na escolha dos autovetores correspondentes (Golyandina et al., 2001).

A SVD em (4.3) é ótima, no sentido de que entre todas as matrizes $\mathbf{X}^{(p)}$ de posto inferior a d ($p < d$), a matriz dada por $\sum_{i=1}^p \mathbf{X}_i$ fornece a melhor aproximação da matriz trajetória \mathbf{X} , de modo o mínimo é atingido através da norma $\|\mathbf{X} - \mathbf{X}^{(p)}\|$. Note-se que $\|\mathbf{X}\|^2 = \sum_{i=1}^d \lambda_i$ e $\|\mathbf{X}_i\|^2 = \lambda_i$ para $i = 1, \dots, d$. Assim que, de forma semelhante à análise de componentes principais (ACP), a contribuição de cada matriz elementar \mathbf{X}_i na expansão (4.3) à matriz de trajetória \mathbf{X} pode-se considerar como a proporção $\lambda_i / \sum_{i=1}^d \lambda_i$. Consequentemente, proveniente também da ACP, são as contribuições acumuladas transmitidas pelas p primeiras matrizes elementares ou menos, que podem ser explicadas pela aproximação ótima da matriz trajetória através da proporção $\sum_{i=1}^p \lambda_i / \sum_{i=1}^d \lambda_i$ (Golyandina et al., 2001).

Outra característica ótima refere-se às propriedades das direções determinadas pelos autovetores U_1, \dots, U_d . Especificamente, o primeiro autovetor U_1 determina a direção de forma tal que a variação das projeções dos vectores desfasados nessa direção é

de máxima. Cada autovetor subsequente determina uma direção que é ortogonal a todas as direções anteriores, e a variação das projeções dos vetores desfasados nesta direção este também é máxima. Portanto, é natural chamar a direção do i -ésimo autovetor U_i a i -ésima direção principal. Note-se que as matrizes elementares $\mathbf{X}_i = U_i W_i^T$ são construídas a partir das projeções dos vetores desfasados nas i -ésimas direções. Este ponto de vista sobre a SVD da matriz trajetória composto por vetores L – *defasados* e um apelo às associações com a análise de componentes principais leva a seguinte terminologia: chame-se o vetor U_i a i -ésimo (principal) autovetor, o vetor V_i será chamado o i -ésimo vetor fator, e o vetor $W_i = \sqrt{\lambda_i} V_i$ o vetor da i -ésima componente principal.

Para aplicar a *Decomposição em Valores Singulares* na matriz de trajetória \mathbf{X} do exemplo proporcionado acima, o primeiro passo é o cálculo de \mathbf{S} , e dos autovalores e autovetores de \mathbf{S} :

$$\mathbf{S} = \mathbf{X}\mathbf{X}^T = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 30 & 40 & 50 \\ 40 & 54 & 68 \\ 50 & 68 & 86 \end{bmatrix}$$

Autovalores \mathbf{S} :

$$\lambda_1 \cong 169,29; \lambda_2 \cong 0,71 \text{ e } \lambda_3 \cong 0$$

Autovetores de \mathbf{S} :

$$U_1 = \begin{bmatrix} -0,418 \\ -0,565 \\ -0,712 \end{bmatrix}, \quad U_2 = \begin{bmatrix} 0,812 \\ 0,120 \\ -0,572 \end{bmatrix}, \text{ e} \quad U_3 = \begin{bmatrix} 0,408 \\ -0,816 \\ 0,408 \end{bmatrix}$$

Considerando que $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ e os resultados dos autovalores, o posto da matriz de trajetória \mathbf{X} é $d = 2$, dado que $\lambda_3 \cong 0$. Por tanto, a decomposição em valores singulares de \mathbf{X} por ser escrita como:

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \end{bmatrix} = \mathbf{X}_1 + \mathbf{X}_2$$

Lembre-se que cada \mathbf{X}_i é o resultado de fazer $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$ e $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$, neste caso do exemplo, $\mathbf{X}_1 = \sqrt{\lambda_1} U_1 V_1^T$ e $V_1 = \mathbf{X}^T U_1 / \sqrt{\lambda_1}$; $\mathbf{X}_2 = \sqrt{\lambda_2} U_2 V_2^T$ e $V_2 = \mathbf{X}^T U_2 / \sqrt{\lambda_2}$. Fazendo o Cálculo de V_i obtém-se como resultado:

$$V_1 = \frac{\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} -0,418 \\ -0,565 \\ -0,712 \end{bmatrix}}{\sqrt{169,29}} = \begin{bmatrix} -0,283 \\ -0,413 \\ -0,543 \\ -0,674 \end{bmatrix}$$

$$V_2 = \frac{\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 0,812 \\ 0,120 \\ -0,572 \end{bmatrix}}{\sqrt{0,71}} = \begin{bmatrix} -0,788 \\ -0,361 \\ 0,066 \\ 0,494 \end{bmatrix}$$

Com o resultado de V_1 e V_2 calcula-se posteriormente cada \mathbf{X}_i assim:

$$\begin{aligned} \mathbf{X}_1 &= \sqrt{169,29} \begin{bmatrix} -0,418 \\ -0,565 \\ -0,712 \end{bmatrix} \begin{bmatrix} -0,283 & -0,413 & -0,543 & -0,674 \end{bmatrix} \\ &= \begin{bmatrix} 1,539 & 2,247 & 2,953 & 3,661 \\ 2,079 & 3,036 & 3,993 & 4,950 \\ 2,621 & 3,827 & 5,033 & 6,239 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \mathbf{X}_2 &= \sqrt{0,71} \begin{bmatrix} 0,812 \\ 0,120 \\ -0,572 \end{bmatrix} \begin{bmatrix} -0,788 & -0,361 & 0,066 & 0,494 \end{bmatrix} \\ &= \begin{bmatrix} -0,539 & -0,247 & 0,047 & 0,339 \\ -0,079 & -0,036 & 0,007 & 0,050 \\ 0,379 & 0,173 & -0,033 & -0,239 \end{bmatrix} \end{aligned}$$

Logo

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2$$

$$\begin{aligned} \mathbf{X} &= \begin{bmatrix} 1,539 & 2,247 & 2,953 & 3,661 \\ 2,079 & 3,036 & 3,993 & 4,950 \\ 2,621 & 3,827 & 5,033 & 6,239 \end{bmatrix} + \begin{bmatrix} -0,539 & -0,247 & 0,047 & 0,339 \\ -0,079 & -0,036 & 0,007 & 0,050 \\ 0,379 & 0,173 & -0,033 & -0,239 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \end{bmatrix} \end{aligned}$$

4.1.2.

Reconstrução

As próximas duas fases: *agrupamento* e *média diagonal* formam o estágio de *reconstrução*. A fase de *agrupamento* corresponde à divisão das matrizes, calculadas na fase de SVD, em vários grupos e somar ditas matrizes dentro de cada grupo, isto é, a

concentração dos elementos com características semelhantes ou com maior correlação entre si, com o intuito de organizá-las em um mesmo grupo. O resultado desta fase é uma representação da matriz trajetória como a soma de várias matrizes resultantes.

A última fase chamada *média diagonal*, transforma cada matriz resultante em uma série temporal, a qual é uma componente aditiva da série inicial. Esta é uma operação linear e mapeia a matriz trajetória da série inicial na própria série inicial, resultando assim, na nova formação da série temporal.

4.1.2.1.

Agrupamento das Autotriplas

A fase de *agrupamento* tem como propósito a distinção das componentes aditivas da série temporal em termos de matrizes separáveis, ou o que corresponde à divisão das matrizes elementares \mathbf{X}_i em vários grupos, para depois assim somar as matrizes dentro de cada grupo. Desta forma, uma vez que a expressão (4.3) é obtida, o procedimento de *agrupamento* decompõe o conjunto de índices $\{1, \dots, d\}$ em m subconjuntos disjuntos $\{I_1, \dots, I_m\}$ utilizando algum método de agrupamento como pode ser por exemplo: *Análise Gráfica dos Valores Singulares*, *Análise Gráfica de Vetores Singulares*, *Análise Gráfica da matriz de correlações ponderadas* ou *Análise de Clusters Usando Métodos Hierárquicos Aglomerativos* (Um detalhe destas técnicas será apresentado na Seção 4.4 e no Apêndice A. Adicionalmente, em Cassiano (2014) são analisadas em profundidade técnicas de agrupamento baseadas em densidade); e posteriormente, ditos subconjuntos (matrizes elementares) são somados dentro de cada grupo, agrupando os elementos que contêm maior grau de correlação, em grupos homogêneos. Desta forma, serão selecionados pelo usuário r das d autotriplas.

Seja $I = \{i_1, \dots, i_r\}$ um grupo de índices. Então, a matriz \mathbf{X}_I correspondente ao grupo I é definida como $\mathbf{X}_I = \mathbf{X}_{i_1} + \mathbf{X}_{i_2} + \dots + \mathbf{X}_{i_r}$. A divisão do conjunto de índices $1, \dots, d$ nos subconjuntos disjuntos I_1, \dots, I_m corresponde à representação:

$$\mathbf{X} = \mathbf{X}_{I_1} + \mathbf{X}_{I_2} + \dots + \mathbf{X}_{I_m} \quad (4.4)$$

\mathbf{X}_I está relacionado com o " sinal " de Y_T , enquanto o restante da subtração $(d-r)$ indica o termo de erro das autotriplas ε .

O procedimento de escolher os conjuntos I_1, \dots, I_m é chamado de agrupamento *das autotriplas*, este fornece o último parâmetro da técnica SSA. A escolha deste parâmetro utiliza o conceito de separabilidade que é apresentado na Seção 4.3.

Ao aplicar a fase do agrupamento das autotriplas para o exemplo da série temporal $Y = [1, 2, 3, 4, 5, 6]$, note-se que na fase de SVD, obtiveram-se apenas dois componentes diferentes de zero e; portanto, nesta fase podem-se definir no máximo dois grupos: $I_1 = \{1\}$, $I_2 = \{2\}$.

4.1.2.2.

Média Diagonal Secundária

Nesta última fase, o grupo de m componentes selecionadas na fase anterior é usado para reconstruir as componentes determinísticas da série temporal. A ideia básica da *média diagonal* é transformar cada matriz \mathbf{X}_I da decomposição agrupada (4.4) em uma matriz Hankel que pode ser posteriormente convertida em uma série temporal, a qual é uma componente aditiva da série inicial Y_T , o que significa transformar cada um dos termos $\mathbf{X}_{I_1}, \mathbf{X}_{I_2}, \dots, \mathbf{X}_{I_m}$ nas séries reconstruídas $\mathbf{y}_{I_1}, \mathbf{y}_{I_2}, \dots, \mathbf{y}_{I_m}$ através do processo de Hankelização ou *média diagonal*.

Ao aplicar o operador linear de Hankelização (\mathcal{H}) obtém-se a transformação das matrizes resultantes \mathbf{X}_I em séries, visto que ele atua em uma matriz arbitrária \mathbf{Z} , por exemplo, transformando-a em uma matriz Hankel, que conseqüentemente é uma matriz trajetória e posteriormente em uma série. Em outras palavras, se z_{ij} representa um elemento da matriz arbitrária \mathbf{Z} , então, k -ésimo termo da série histórica reconstruída é obtido através do cálculo da média de z_{ij} sobre todos os i, j tal que $i + j = k + 1$. Este procedimento é chamado de *média diagonal* ou *Hankelização da matriz \mathbf{Z}* . Desta forma, no contexto da segunda fase do estágio de reconstrução, o operador \mathcal{H} contabiliza cada uma das entradas ao longo das linhas paralelas às diagonais secundárias das matrizes \mathbf{X}_I como sendo as médias. O resultado da Hankelização da matriz \mathbf{Z} é a matriz de Hankel $\mathcal{H}\mathbf{Z}$, a qual como foi dito anteriormente, é a matriz de trajetória correspondente à série obtida como resultado do cálculo da *média diagonal* (Golyandina et al., 2001).

De acordo com Golyandina e Zhigljavsky (2013), considerando-se \mathbf{X} uma matriz trajetória com elementos $x_{ij}, 1 \leq i \leq L, 1 \leq j \leq K$. Admita-se que $L^* = \min(L, K)$, $K^* = \max(L, K)$ e $T = L + K - 1$. Além disso, $x_{ij}^* = x_{ij}$, se $L < K$ e $x_{ij}^* = x_{ji}$, caso contrário.

Com a utilização deste procedimento, a matriz \mathbf{X} transforma-se na série y_1, \dots, y_T da seguinte maneira:

$$y_k = \begin{cases} \frac{1}{k} \sum_{m=1}^k x_{m,k-m+1}^* & \text{para } 1 \leq k < L^*, \\ \frac{1}{L^*} \sum_{m=1}^{L^*} x_{m,k-m+1}^* & \text{para } L^* \leq k \leq K^*, \\ \frac{1}{T-k+1} \sum_{m=k-K^*+1}^{T-K^*+1} x_{m,k-m+1}^* & \text{para } K^* < k \leq T. \end{cases} \quad (4.5)$$

Isto corresponde a uma média dos elementos da matriz sobre as “antidiagonais” ou diagonais paralelas à diagonal secundária $i + j = k + 1$, com $k = 1, 2, \dots, T$.

Por exemplo, escolhendo $k = 1$, tem-se:

$$y_1 = y_{1,1}$$

para $k = 2$, tem-se:

$$y_2 = \frac{(y_{1,2} + y_{2,1})}{2}$$

para $k = 3$, obtém-se

$$y_3 = \frac{(y_{1,3} + y_{3,1} + y_{2,2})}{3}$$

e, assim por diante até a obtenção do último termo da série resultante.

Quando a matriz \mathbf{Z} possui um número maior de linhas do que colunas ($L > K$), o processo para a consecução dos elementos da matriz $\mathcal{H}\mathbf{Z}$ é similar ao caso anterior, no qual $L \leq K$. As alterações referem-se à substituição $L \leftrightarrow K$ e à transposição dos elementos de \mathbf{Z} ($z_{ij} \rightarrow z_{ji}$).

Na prática, a matriz trajetória \mathbf{X} da série original possui um número de linhas inferior ao de colunas ($L < K$). Isto porque, como T é fixo e existem limites para L ($2 \leq L$

$\leq T/2$), o valor mínimo de K fica determinado por $K_{min} = T - L_{max} + 1 = T/2 + 1$; onde o L_{max} denota o valor máximo de L . Desta forma tem-se que sempre $L_{max} < K_{min}$.

Ao aplicar o procedimento Hankelização a todos os componentes da matriz de (4.4), obtém-se outra expansão:

$$\mathbf{X} = \tilde{\mathbf{X}}_{I_1} + \tilde{\mathbf{X}}_{I_2} + \dots + \tilde{\mathbf{X}}_{I_m} \quad (4.6)$$

sendo $\tilde{\mathbf{X}}_{I_1} = \mathcal{H}\mathbf{X}_{I_1}$. Isto é equivalente à decomposição da série original $Y_T = (y_1, y_2, \dots, y_T)$ em uma soma de m séries reconstruídas:

$$Y_T = \tilde{Y}_T^{(1)} + \tilde{Y}_T^{(2)} + \dots + \tilde{Y}_T^{(m)} = \sum_{k=1}^m \tilde{Y}_T^{(k)} \quad (4.7)$$

Com o intuito de exemplificar esta última fase foram calculadas as médias dos elementos cujos índices quando somados obtém-se os mesmos valores, da seguinte forma:

$$\mathbf{X}_1 = \begin{bmatrix} 1,539 & 2,247 & 2,953 & 3,661 \\ 2,079 & 3,036 & 3,993 & 4,950 \\ 2,621 & 3,827 & 5,033 & 6,239 \end{bmatrix}$$

$L = 3$, $K = 4$ e adicionalmente, $L^* = \min(L, K) = \min(3, 4) = 3$ e $K^* = \max(L, K) = \max(3, 4) = 4$.

- Para $k = 1$ escolhe-se a primeira expressão da Equação (4.5) já que cumpre com a restrição $1 \leq k < L^*$. Desta forma, temos que:

$$y_1 = \frac{x_{1,1}}{1} = x_{1,1} = 1,538$$

- Para $k = 2$ escolhe-se a primeira expressão da Equação (4.5) já que continua cumprindo com a restrição $1 \leq k < L^*$. Desta forma, temos que:

$$y_2 = \frac{(x_{1,2} + x_{2,1})}{2} = 2,162$$

- Para $k = 3$ escolhe-se a segunda expressão da Equação (4.5) já que cumpre com a restrição $L^* \leq k \leq K^*$. Desta forma, temos que:

$$y_3 = \frac{(x_{1,3} + x_{2,2} + x_{3,1})}{3} = 2,870$$

- Para $k = 4$ escolhe-se a segunda expressão da Equação (4.5) já que cumpre com a restrição $L^* \leq k \leq K^*$. Desta forma, temos que:

$$y_4 = \frac{(x_{1,4} + x_{2,3} + x_{3,2} + x_{4,1})}{3} = 3,826$$

- Para $k = 5$ escolhe-se a terceira expressão da Equação (4.5) já que cumpre com a restrição $K^* < k \leq T$. Desta forma, temos que:

$$y_5 = \frac{(x_{2,4} + x_{3,2})}{2} = 4,991$$

- Para $k = 6$ escolhe-se a terceira expressão da Equação (4.5) já que cumpre com a restrição $K^* < k \leq T$. Desta forma, temos que:

$$y_6 = \frac{x_{3,4}}{1} = 6,239$$

Logo

$$\tilde{\mathbf{X}}_1 = \begin{bmatrix} 1,538 & 2,162 & 2,870 & 3,827 \\ 2,162 & 2,870 & 3,827 & 4,991 \\ 2,870 & 3,827 & 4,991 & 6,239 \end{bmatrix}$$

De forma similar é feito cálculo da média diagonal para \mathbf{X}_2 :

$$\mathbf{X}_2 = \begin{bmatrix} -0,538 & -0,246 & 0,047 & 0,339 \\ -0,080 & -0,036 & 0,007 & 0,050 \\ 0,379 & 0,173 & -0,033 & -0,239 \end{bmatrix}$$

Obtendo como resultado:

$$\tilde{\mathbf{X}}_2 = \begin{bmatrix} -0,538 & -0,162 & 0,130 & 0,173 \\ -0,162 & 0,130 & 0,173 & 0,009 \\ 0,130 & 0,173 & 0,009 & -0,239 \end{bmatrix}$$

Desta forma, a série reconstruída é a seguinte:

$$\mathbf{X} = \tilde{\mathbf{X}}_1 + \tilde{\mathbf{X}}_2 = \begin{bmatrix} 1,538 \\ 2,162 \\ 2,870 \\ 3,827 \\ 4,991 \\ 6,239 \end{bmatrix} + \begin{bmatrix} -0,538 \\ -0,162 \\ 0,130 \\ 0,173 \\ 0,009 \\ -0,239 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}$$

Observe que tanto $\tilde{\mathbf{X}}_1$ quanto $\tilde{\mathbf{X}}_2$ são matrizes Hankel e, portanto, são a matrizes de trajetória correspondentes às séries obtidas como resultado do cálculo da *média diagonal*.

4.1.3.

Previsão

Neste último estágio são realizadas as previsões utilizando SSA. No entanto, uma previsão só pode ser feita se um modelo foi construído. O modelo deve ser derivado a partir dos dados ou pelo menos verificado contra os dados. Estes modelos em SSA podem ser descritos através das *fórmulas recorrentes lineares (LRFs - Linear Recurrent Formula)*. Para obter uma previsão via SSA o requisito básico é que a série $Y_T = (y_1, y_2, \dots, y_T)$ satisfaça as *LRFs* de ordem ou dimensão $L - 1$, para as quais é necessário que existam os coeficientes a_1, \dots, a_{L-1} tais que:

$$Y_{T-i} = \sum_{k=1}^{L-1} a_k Y_{T-i-k} \quad (4.8)$$

com $0 \leq i \leq T - L$ e $a_k \neq 0$

Os coeficientes a_1, \dots, a_{L-1} podem ser determinados através dos autovetores obtidos na fase da SVD. Para um maior detalhe do procedimento, consultar em Golyandina et al. (2001).

A classe de séries temporais governadas por *LRFs* é bastante abrangente e importante para aplicações práticas. Desta forma, uma série temporal é governada por *LRFs* se e somente se ela pode ser representada como uma combinação linear dos produtos de séries exponenciais, polinomiais e harmônicas (Golyandina & Zhigljavsky, 2013) da seguinte forma:

$$Y_T = \sum_{p=1}^q a_p(T) e^{T\mu_p} \sin(2\pi T\omega_p + \varphi_p) \quad (4.9)$$

sendo $a_p(T)$ a representação dos polinômios; μ_p , ω_p e φ_p são constantes arbitrárias. E q denota o número de termos literalmente independentes.

Em Golyandina et al. (2001), a ideia das *fórmulas recorrentes lineares* são apresentadas expressando séries governadas por uma *LRF* de ordem mínima, cuja dimensão é representada por d . Segundo as características de decomposição de SSA

(Seção 4.1.1.2) existem no máximo d valores singulares no nulos na SVD da matriz \mathbf{S} ; e portanto, mesmo que L e K sejam maiores do que d , precisa-se de no máximo d matrizes elementares \mathbf{X}_i para reconstruir a série.

Apesar disso, na prática, dificilmente espera-se ter uma série que seja governada por uma LRF de dimensão pequena. Ao contar com uma série temporal que satisfaça uma LRF (Eq. 4.8) pode-se dar a continuidade desta para um número arbitrário de passos à frente usando a mesma LRF . É importante que qualquer LRF que governe uma determinada série temporal proporcione a mesma continuação da série original Y_T , o que faz o que não necessariamente se precise uma LRF de dimensão mínima. Esta continuação é obtida ao aplicar uma LRF nos termos passados da série Y_T .

O autor Danilov (1997a, 1997b), proporcionou um resultado fundamental para a previsão com SSA, no qual se formula o seguinte: se a dimensão r do espaço linear \mathcal{Q}_r , também conhecido como espaço trajetória, gerado pelas colunas da matriz de trajetória é menor do que o comprimento da janela L e esse espaço é um espaço “*não-vertical*”, então a série satisfaz uma LRF natural de dimensão $L - 1$. Em forma geral, diz-se que \mathcal{Q}_r é um espaço “*não-vertical*” se a base $e_L \notin \mathcal{Q}_r$, sendo $e_L = (0, 0, \dots, 1)^T \in \mathbb{R}^L$ (Golyandina et al., 2001).

Sob a suposição de espaço “*não-vertical*”, o espaço trajetória \mathcal{Q}_r produz $LRFs$ adequadas, as quais podem ser usadas para previsão das componentes da série, correspondendo às matrizes de posto unitário selecionadas no segundo estágio da técnica SSA; o que representa a seleção de um espaço r – *dimensional* ($\mathcal{Q}_r \in \mathbb{R}^L$) gerado pelos correspondentes *autovetores da esquerda*. Este processo de previsão é conhecido como *previsão recorrente*.

Por outro lado, é importante destacar que existem outras metodologias de previsão via SSA tais como: *previsão vetorial* e *previsão simultânea*; no entanto, neste trabalho será considerado o algoritmo univariado de *previsão recorrente* baseado no SSA básico (Introduzida por Danilov, 1997a, 1997b) descrita na seguinte Seção.

4.1.3.1.

Algoritmo de Previsão Recorrente

A principal suposição a ser assumida é que o comprimento da janela L foi escolhido de forma a separar seguramente o sinal do ruído. Neste sentido, podem-se fazer previsões utilizando-se as autotriplas escolhidas (Golyandina e Zhigljavsky, 2013).

A previsão dos valores futuros \hat{y}_{T+h} , com $h = 1, \dots, M$, é obtida a partir da seguinte expressão recorrente:

$$\hat{y}_i = \sum_{j=1}^{L-1} a_j \hat{y}_{i-j}; \quad i = T + 1, \dots, T + M \quad (4.10)$$

Sendo a_j os coeficientes da combinação linear entre os $L - 1$ últimos termos da série considerada, ou seja:

$$\hat{y}_{T+h} = \sum_{j=1}^{L-1} a_j \hat{y}_{(T+h)-j}; \quad h = 1, \dots, M \quad (4.11)$$

Os coeficientes a_j podem ser representados de forma de vetorial: $R = (a_1, \dots, a_{L-1})^T$. O vetor R pode ser expresso como segue:

$$R = \frac{1}{1 - v^2} \sum_{j=1}^r \pi_j U_j^{\nabla}, \quad (4.12)$$

Segundo Mahmoudvand et al. (2013), assume-se que o vetor U_j^{∇} denota as primeiras $L - 1$ componentes do autovetor U_j e que π_j é a última componente de U_j ($j = 1, \dots, r$). O termo $v^2 = \sum_{j=1}^r \pi_j^2$.

Considerando as notações anteriores, as previsões pelo algoritmo *recorrente* $(\hat{y}_{T+1}, \dots, \hat{y}_{T+M})$ podem ser obtidas por meio da seguinte fórmula:

$$\hat{y}_i = \begin{cases} \tilde{X}_i, & i = 1, \dots, T \\ \sum_{j=1}^{L-1} a_j \hat{y}_{i-j}, & i = T + 1, \dots, T + M \end{cases} \quad (4.13)$$

Os valores $\tilde{X}_1, \dots, \tilde{X}_T$ são os valores da série reconstruída obtidos em (4.7), enquanto que $\hat{y}_{T+1}, \dots, \hat{y}_{T+M}$ formam os M termos da *previsão recorrente SSA*. A expressão pode ser rescrita como segue:

$$\hat{y}_i = \begin{cases} \tilde{X}_i, & i = 1, \dots, T \\ R^T Q_i, & i = T + 1, \dots, T + M \end{cases} \quad (4.14)$$

onde, $Q_i = [\hat{y}_{i-L+1}, \dots, \hat{y}_{i-1}]^T$.

Exemplificando este estágio, lembre-se que o vetor U_j^V contém os $L - 1$ elementos do vetor U_j , com $j = 1, \dots, r$. Considerando-se $r = 1$ e $M = 1$, tem-se:

$$U_1^V = \begin{bmatrix} -0,418 \\ -0,565 \end{bmatrix}$$

O valor de π_j corresponde ao último elemento do vetor U_j .

$$\pi_1 = -0,712$$

O valor de v^2 é dado pela seguinte fórmula:

$$v^2 = \sum_{j=1}^r \pi_j^2 = \pi_1^2 = (-0,712)^2 = 0,5069$$

O vetor R é obtido por:

$$R = \frac{1}{1 - v^2} \sum_{j=1}^r \pi_j U_j^V = \frac{1}{1 - 0,5069} (\pi_1 U_1^V) = 2,028 \left(-0,712 \begin{bmatrix} -0,418 \\ -0,565 \end{bmatrix} \right) = 2,028 \begin{bmatrix} 0,298 \\ 0,402 \end{bmatrix}$$

$$R = \begin{bmatrix} 0,604 \\ 0,815 \end{bmatrix}$$

Para efetuar a previsão é utilizada Equação (4.14). Assim, a previsão \hat{y}_7 é dada por:

$$Q_7 = [\hat{y}_{7-3+1}, \dots, \hat{y}_{7-1}]^T = [\hat{y}_5, \hat{y}_6]^T = [4,991 \quad 6,239]^T$$

Obs: \hat{y}_5 e \hat{y}_6 correspondem à posição 5 e 6 respectivamente da série \tilde{X}_1 , exemplificada na parte final da fase 2 do estágio 2 de agrupamento. Desta forma R e Q_i produzem a primeira previsão um passo à frente para $\hat{y}_7 = 0,815\hat{y}_6 + 0,604\hat{y}_5$:

$$\hat{y}_7 = R^T Q_7 = [0,604 \quad 0,815] \begin{bmatrix} 4,991 \\ 6,239 \end{bmatrix} = 3,01 + 5,08 = 8,09$$

E assim por diante para os M valores a serem previstos. Como neste exemplo $M = 1$, a previsão finaliza com o cálculo de \hat{y}_7 .

4.2.

Tamanho do Parâmetro L de Defasagem

A determinação do parâmetro L de defasagem (conhecido também como tamanho da janela) é de suma importância em SSA, pois como foi dito anteriormente, L é o único parâmetro do estágio de decomposição e o principal parâmetro da metodologia SSA, repercutindo assim nas fases subsequentes, já que a escolha do comprimento da janela L é um aspecto crítico que se é efetuada de forma inadequada implicaria que nenhum agrupamento conduziria a uma boa decomposição SSA. É importante ressaltar que uma adequada seleção de L depende do problema de interesse e da informação preliminar acerca da série temporal em questão. Não existe uma regra geral para a escolha do comprimento da janela, daí que a principal dificuldade que se pode apresentar como resultado das variações de L é influenciar ambas as características de separabilidade de SSA: forte e fraca, ou seja, a ortogonalidade das subséries apropriadas da série original e a proximidade dos valores singulares, ou seja, dificulta o bom agrupamento e, por conseguinte, diminui a precisão na aproximação da série. Resultados teóricos estabelecem que L deveria ser o suficientemente grande, mas não superior a $T/2$, tornando a decomposição da série temporal mais detalhada na medida em que o comprimento da janela aumenta.

Um aspecto que deve ser tratado com muita atenção é quando se trabalha com séries temporais com componentes periódicas (séries sazonais). Portanto, se com antecedência sabe-se que a série temporal é uma série sazonal, então para conseguir a separabilidade desta componente periódica é recomendável tomar o tamanho da janela L proporcional ao período da sazonalidade dos dados (Golyandina et al., 2013).

Para exemplificar o efeito que tem a escolha do tamanho da janela L , Golyandina et al. (2001) apresenta o resultado da decomposição de uma série caracterizada como complexa. A série utilizada faz referência a dados diários ao longo de 14 anos, conduzindo a um total de 5113 observações. A série possui dois comportamentos periódicos, periodicidade semanal (período $T_1 = 7$) e anual ($T_2 = 365$). Assim, desde que T_2 seja muito maior do que T_1 , resulta natural considerar L múltiplo de T_2 .

A escolha de $L = T_2$ garantiu a extração simultânea da periodicidade anual e semanal. Além disso, este valor de L permitiu a extração da tendência da série usando

uma única autotripla principal. Enquanto que, se o L for aumentado em três vezes ($L = 3 \times T_2 = 1095$), a componente da tendência fica misturada com as componentes de periodicidade. Então, em geral, valores altos para L conduzem à violação da suposição de separabilidade forte. Golyandina et al. (2001) salienta que, se não é importante o problema da separação da tendência (ou nível) e periodicidade, então valores de L superiores a T_2 também podem ser utilizados; e neste caso, tem-se a vantagem de uma decomposição mais detalhada.

Ao utilizar $L = 365$, a estrutura geral da série é recolhida com apenas 5 autotriplas, enquanto que, ao considerar $L = 1095$ precisa-se de dezenas de autotriplas. Por outro lado, se for considerado um L pequeno (em relação ao período anual, por exemplo, $L = 28$), então o comportamento global da série deve ser descrito por apenas uma autotripla principal. Isto indica que, em situações semelhantes, ao aumentar o comprimento da janela, há a necessidade de considerar mais autotriplas na etapa de reconstrução da série, para que estas possam captar melhor as características das componentes da série.

O foco de diversos artigos centra-se neste tópico gerando como resultado uma série de estudos (Golyandina & Vlassieva, 2009; Golyandina, 2010; Korobeynikov, 2010), que embora contribuam na escolha do parâmetro L , não fornecem uma proposta metodológica formal fechada que não dependa da tipologia dos dados nem do problema em questão.

Considerando a situação atual para seleção do tamanho da janela L , esta tese baseia-se em uma escolha sequencial sugerida por Golyandina & Korobeynikov (2014) caso a série temporal possua uma estrutura complexa (Séries longas). SSA *sequencial* consta de duas fases, na primeira fase, a tendência ou nível é extraído, em seguida na segunda fase, as componentes periódicas são detectadas e extraídas a partir dos residuais da extração da tendência/nível, considerando L proporcional ao período sazonal.

4.3.

Separabilidade

O conceito de *separabilidade* responde à pergunta natural de como escolher parâmetros para construir uma decomposição adequada da série histórica observada e quando isso é possível. Ela caracteriza quão bem as diferentes componentes podem ser separadas umas das outras. A decomposição da série Y_T só poderá ter êxito se as componentes aditivas resultantes da série são (aproximadamente) separáveis umas das outras. Desta forma as fases de SVD e *agrupamento das autotriplas* apoiam-se nesta propriedade denominada separabilidade.

Considere a série original Y_T , suponha que ela é obtida através da soma de duas subséries ($Y_T = Y_T^{(1)} + Y_T^{(2)}$). Nesta situação, a separabilidade das subséries $Y_T^{(1)}$ e $Y_T^{(2)}$ implica que os termos na SVD da matriz trajetória \mathbf{X} podem ser divididos em dois diferentes grupos, de forma que as somas dos termos dentro de cada grupo proporciona as matrizes trajetórias $\mathbf{X}^{(1)}$ e $\mathbf{X}^{(2)}$ das subséries $Y_T^{(1)}$ e $Y_T^{(2)}$, respectivamente.

A separabilidade implica, em termos teóricos, que cada linha da matriz trajetória $\mathbf{X}^{(1)}$ da primeira subsérie é ortogonal a cada linha da matriz trajetória $\mathbf{X}^{(2)}$ da segunda subsérie, e o mesmo vale para as colunas. Caso esta suposição de ortogonalidade se cumpra, pode-se dizer que as subséries $Y_T^{(1)}$ e $Y_T^{(2)}$ são *fracamente separáveis*. Existe uma noção de separabilidade mais restrita (e mais desejável na prática), conhecida como *separabilidade forte*, a qual além da suposição de separabilidade fraca, impõe restrições sobre os valores singulares das matrizes trajetórias $\mathbf{X}^{(1)}$ e $\mathbf{X}^{(2)}$, considerando que os conjuntos formados pelos valores singulares são disjuntos. Não entanto, segundo Golyandina et al. (2001), na prática, não acontece *separabilidade exata*. Nas situações práticas, fala-se somente de *separabilidade aproximada*. Comumente, avalia-se a qualidade da separabilidade através de uma medida natural de dependência entre as subséries. Essa medida é denominada como “*correlação ponderada*” ou “*w-correlation*”.

Ao se referir à *separabilidade exata*, fala-se sobre a ortogonalidade das linhas e colunas das matrizes trajetória $\mathbf{X}^{(1)}$ e $\mathbf{X}^{(2)}$, que em termos de Álgebra Linear significa que todos os produtos internos das linhas e colunas são zero. Na linguagem estatística, consiste em que todas as covariâncias (E, portanto, as correlações – Cossenos dos ângulos formados entre os vetores correspondentes) são zero.

Por conseguinte, este fato pode ser considerado como uma característica da separabilidade das duas séries $Y_T^{(1)}$ e $Y_T^{(2)}$, sendo $\rho^{(L,K)}$ o *coeficiente de correlação máxima*, que é o valor absoluto máximo das correlações entre as linhas e entre as colunas das matrizes trajetórias destas duas séries. Desta forma, dirá-se que as duas séries $Y_T^{(1)}$ e $Y_T^{(2)}$ são aproximadamente separáveis se todas as correlações entre as linhas e as colunas das matrizes trajetórias $\mathbf{X}^{(1)}$ e $\mathbf{X}^{(2)}$ estão perto de zero.

A seguinte expressão chamada de *correlação ponderada* (*w-correlation*) é uma medida natural que quantifica a dependência linear (desvio) entre duas séries $Y_T^{(1)}$ e $Y_T^{(2)}$:

$$\rho_{12}^{(\omega)} = \frac{\left(Y_T^{(1)}, Y_T^{(2)} \right)_{\omega}}{\| Y_T^{(1)} \|_{\omega} \| Y_T^{(2)} \|_{\omega}}. \quad (4.15)$$

Sendo $\| \cdot \|$ é a norma euclidiana, $(\cdot)_{\omega}$ é o produto interno. Assim, a norma da i -ésima subsérie é dada por:

$$\|Y_T^{(i)}\|_{\omega} = \sqrt{(Y_T^{(i)}, Y_T^{(i)})_{\omega}}; \quad (4.16)$$

o produto interno entre este par de subséries é:

$$(Y_T^{(i)}, Y_T^{(j)})_{\omega} = \sum_{k=1}^T \omega_k Y_k^{(i)} Y_k^{(j)}, \quad (4.17)$$

$$(i, j = 1, 2);$$

Os pesos ω_k são iguais ao número de vezes que o elemento Y_k aparece na matriz de trajetória \mathbf{X} , para os quais:

$$\omega_k = \begin{cases} i & \text{para } 0 \leq i \leq L^*, \\ L^* & \text{para } L^* \leq i \leq K^*, \\ T - i + 1 & \text{para } K^* < i \leq T. \end{cases} \quad (4.18)$$

Caso valor absoluto da *correlação ponderada* seja pequeno, isto indica que as duas séries são classificadas como *w-ortogonais* ou quase *w-ortogonais*, enquanto que se a correlação ponderada fosse grande, então as séries são classificadas como mal separadas.

Em seções posteriores será apresentada a *correlação ponderada* como uma técnica de agrupamento das componentes resultantes da SVD.

4.4.

Informação Complementar

A informação complementar auxilia na identificação das autotriplas mais relevantes para uma adequada extração das componentes de tendência, periodicidade e ruído; assim como também tornar a situação mais clara e ajudar na escolha dos parâmetros dos modelos. A escolha do parâmetro L depende do objetivo da análise e das informações preliminares sobre a série temporal. Uma seleção acertada de L permite atingir uma boa separabilidade das componentes repercutindo em um apropriado agrupamento das autotriplas.

A continuação encontram-se algumas ferramentas de grande utilidade para separar o sinal do ruído. Para um maior detalhe dos tópicos a serem abordados consultar Golyandina et al., (2001) e Hassani (2007).

4.4.1.

Valores Singulares

Os *valores singulares* são obtidos através da SVD na segunda fase do estágio de *Decomposição* e seu comportamento pode ser observado através do espectro dos valores singulares da matriz de trajetória \mathbf{X} , isto é, um gráfico no qual são apresentados de forma decrescente em magnitude os seus L valores $(\sqrt{\lambda_1} \geq \sqrt{\lambda_2} \geq \dots \geq \sqrt{\lambda_L})$.

Uma das características importantes a ser observada é que o primeiro valor singular, aquele com maior valor absoluto, corresponde à componente de tendência da série. Enquanto que, planícies explícitas (autovalores com magnitudes muito próximas) no espectro dos autovalores indicam os números ordinais, das autotriplas pareadas, associados a uma componente periódica do sinal. Desta forma, podem ser identificados, no gráfico do espectro de \mathbf{X} , os valores singulares pareados cujos valores são quase iguais; e assim consequentemente, detectar as componentes periódicas da série temporal.

Outra visão útil é fornecida através da verificação das quebras no espectro dos valores singulares. Em termos gerais, uma série de ruído puro ou conhecida como puramente residual produz um uma sequência lentamente decrescente dos valores singulares. Se tal ruído é adicionado ao sinal, conformado por algumas autotriplas com valores singulares grandes, em seguida, poder ser observada uma quebra no espectro dos valores singulares de \mathbf{X} , permitindo distinguir as autotriplas que estão associadas ao sinal daquelas que vinculadas ao ruído; em outras palavras, a partir do momento em que se visualiza um decrescimento lento e contínuo dos valores singulares, pode caracterizar o início do ruído da série temporal. Em geral, não existem procedimentos formais que permitam para encontrar tal quebra. Além disso, para sinais complexos e ruídos grandes, as autotriplas do sinal e do ruído podem estar confundidas com relação à ordem de seus valores singulares (Golyandina et al., 2001).

Com o intuito de exemplificar o anteriormente exposto, será analisada a série sobre dos dados mensais de mortes acidentais nos EUA entre 1973 e 1978 ($T = 72$), a qual também abordada por Hassani (2007). A série possui uma tendência linear e uma sazonalidade aditiva como pode ser observado na Figura 4.2.

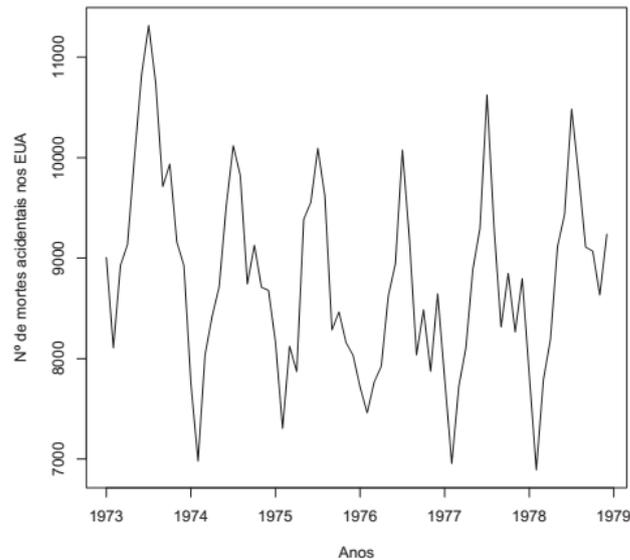


Figura 4.2 – Número mensal de mortes acidentais nos EUA entre 1973 e 1978.

(Fonte: Hassani, 2007)

Na Figura 4.2, pode-se observar a presença de uma periodicidade anual. Assim que seguindo as considerações para a escolha adequada do tamanho da janela e, a fim de obter uma melhor separabilidade desta componente periódica, estabelece-se o comprimento da janela como sendo um valor proporcional ao período 12. Hassani (2007) escolheu $L = 24$, proporcional a esse período. Quanto se utiliza $L = 24$ os autovalores resultantes da SVD da série original podem ser visualizados através espectro dos valores singulares apresentados na Figura 4.3. Note que o primeiro valor singular (autovalor de maior magnitude, posto 1) corresponde à componente de tendência da série.

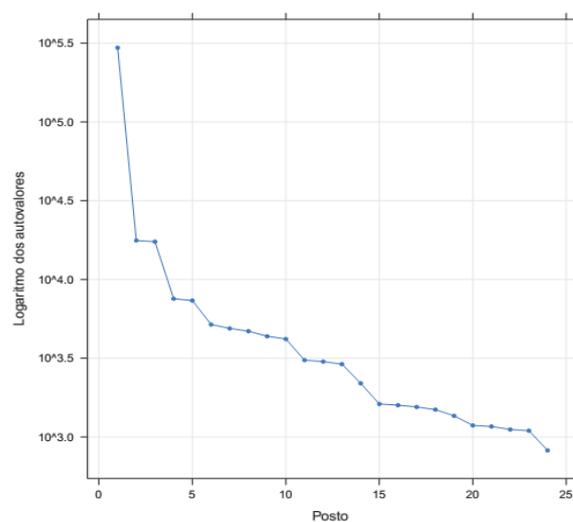


Figura 4.3 – Autovalores resultantes da decomposição da série em 24 componentes.

(Fonte: Hassani, 2007)

Uma vez identificada a componente de tendência, o seguinte a ser feito é determinar as componentes periódicas pareadas. Assim, no espectro dos valores singulares do exemplo, podem-se observar as componentes pareadas periódicas compostas por: (2-3); (4-5) e (11-12), as quais constituem o padrão de sazonalidade da série original começando partir da segunda componente. Existem componentes periódicas que não conseguem ser evidenciadas de forma trivial, as quais correspondem aos autovalores dos postos 6 a 10 e que podem ser também consideradas na formação do comportamento periódico da série.

A partir do posto 15 visualiza-se um decrescimento lento e contínuo dos valores singulares, além de contar com uma quebra na Figura 4.3 (componentes 13 e 14); o que estabelece um limite entre o sinal e o ruído. Resultados que concordam com os resultados teóricos disponíveis na literatura.

4.4.2.

Correlação Ponderada

A *correlação ponderada* é um resultado que se deriva do conceito chamado separabilidade, anteriormente tratado. Estabelece-se que para que exista a separabilidade (aproximada) de duas componentes reconstruídas é necessária que a *correlação ponderada* seja nula ou que esteja próxima de zero, o que indica que as componentes estão altamente correlacionadas e conformam um mesmo grupo de autotriplas. Desta forma, a *correlação ponderada* torna-se uma ferramenta de informação suplementar empregada no estágio de *reconstrução* para o agrupamento da autotriplas.

No processo de agrupamento sugere-se usar a matriz de *correlações ponderadas*, correspondente à decomposição total (nesta decomposição cada autotripla corresponde a uma única componente da matriz da SVD). Este procedimento se fundamenta na análise gráfica da matriz de *correlações ponderadas* (Seção 4.3), e indica as correlações entre as componentes da SVD através de uma escala de cinzas desde o branco até o preto que correspondem aos valores absolutos das correlações de 0 até 1 (Golyandina et al., 2001).

A forma da matriz de *correlações ponderadas* fornece um indicativo de como realizar um agrupamento adequado. Nela espera-se uma distinção bastante clara das componentes em duas partes: a primeira parte está constituída pelas componentes que estejam altamente correlacionadas (tonalidade mais escura), diferenciando o grupo das autotriplas relacionadas ao sinal da série daquelas associadas ao ruído. A segunda parte constitui o ruído, o qual é representado por componentes com correlações baixas (tonalidades claras).

Para exemplificar o uso gráfico da matriz de *correlações ponderadas*, utilizou-se o exemplo da Subseção anterior. A Figura 4.4 mostra as *correlações ponderadas* para 24 componentes reconstruídas em uma grade em escala de cinzas desde o branco até o preto que correspondem aos valores absolutos das correlações de 0 até 1.

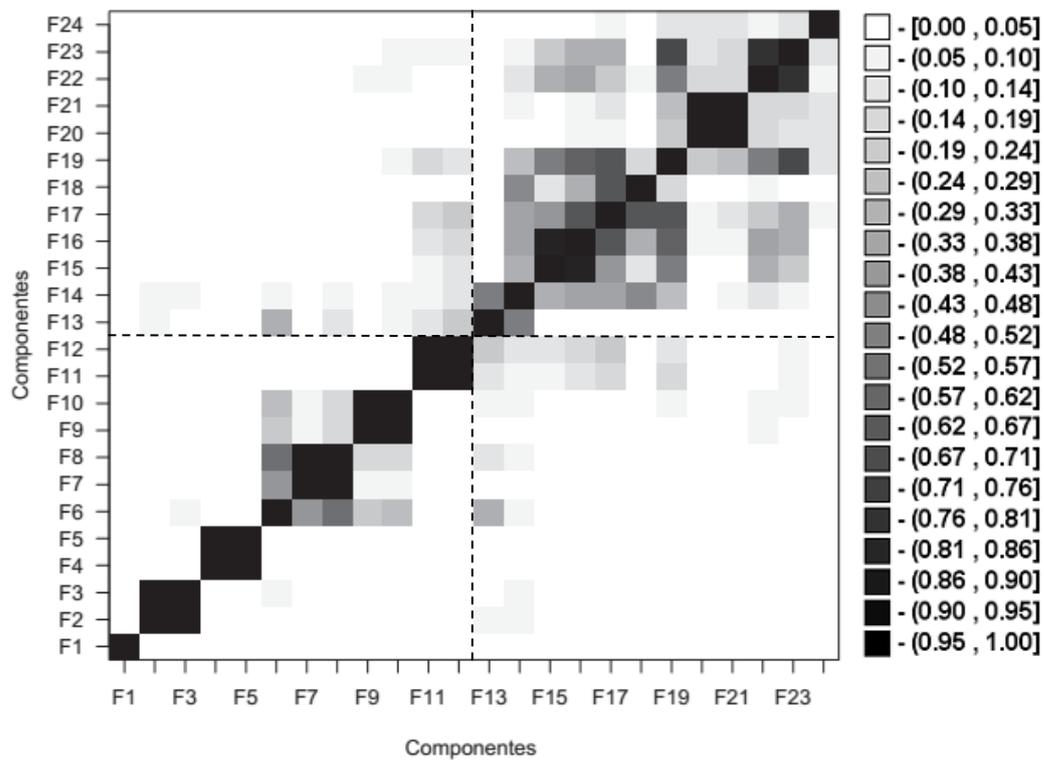


Figura 4.4 – Matriz de *correlações ponderadas* para 24 componentes reconstruídas.

(Fonte: Hassani, 2007)

Nota-se, na Figura 4.3, que o sinal da série parece estar gerado pelas doze autotriplas principais, estabelecendo-se assim que as doze primeiras componentes são suficientes para reconstruir a série original. No entanto, as componentes restantes compõem o ruído, o que significa que a partir da décima terceira componente em diante pode-se apreciar a presença de componentes com correlações menores, representadas através das tonalidades mais claras, fora da diagonal do gráfico.

4.4.3.

Gráficos de Dispersão de Autovetores Pareados

A análise visual dos gráficos de dispersão dos autovetores U_i pareados identifica as componentes harmônicas de uma determinada série, desde que estas componentes sejam separáveis da componente associada o ruído. Ao detectar as componentes de tendência e sazonalidade da série, podem-se separar as componentes oscilatórias daquelas que formam parte da tendência. Assim, com a inspeção visual dos gráficos auxilia-se o procedimento de agrupamento das componentes associadas a este sinal, complementando a informação fornecida pela matriz de *correlações ponderadas* e o espectro dos valores singulares.

Considere um harmônico puro com frequência igual a ω , fase igual a δ , amplitude igual a ξ e período $\rho = \frac{1}{\omega}$ definido como um divisor do tamanho da janela L e K . Quando o parâmetro ρ assume um valor numérico inteiro, então ρ é classificado como *período do harmônico* (Hosseini, 2007). As coordenadas da série temporal de duas componentes ortogonais podem ser dispostas em um diagrama de dispersão.

As funções puras seno e cosseno com frequências, amplitudes e fases iguais resultam em um diagrama de dispersão que exhibe um padrão circular (Morettin & Toloï, 2006). Por outro lado, se $\rho = \frac{1}{\omega}$ é um valor inteiro, logo o diagrama de dispersão exhibe um polígono regular com ρ vértices. Para uma frequência $\omega = \frac{m}{n} < 0.5$ com m e n inteiros e primos, os pontos são os vértices de um gráfico de dispersão de um polígono regular de n vértices (Golyandina et al., 2001). Em vista do anterior, a identificação das componentes que são geradas por um harmônico se reduz à inspeção visual do padrão determinado nos diferentes pares de componentes.

Na Figura 4.5 são apresentados os gráficos de dispersão dos autovetores pareados da série anteriormente usada para exemplificar os conceitos descritos, gráficos que correspondem aos harmônicos com períodos de 12, 6, 4, 3 e 2.5. Eles são ordenados pela sua contribuição (forma) na etapa SVD.

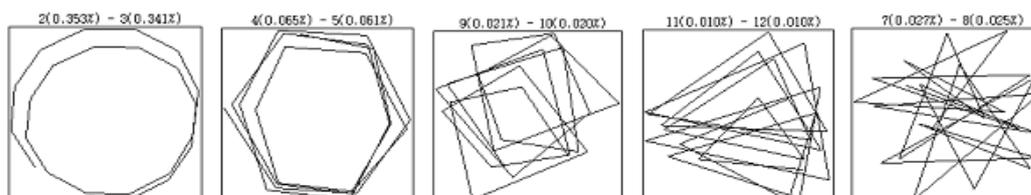


Figura 4.5 – Gráficos de dispersão dos autovetores harmônicos pareados.

(Fonte: Hassani, 2007)

4.4.4.

Análise do Periodograma

A análise do periodograma dos autovetores pode facilitar a identificação de uma componente harmônica geral, indicando qual frequência deve ser considerada; permitindo com isto em fazer um agrupamento adequado. Por conseguinte, devem-se procurar as autotriplas cujas frequências coincidem com as frequências da série original (Hassani, 2007).

Para determinar as frequências das autotriplas é necessário observar se os periodogramas dos autovetores pareados apresentam pontas afiadas ao redor de algumas frequências; caso isto aconteça, as autotriplas que correspondem à dita característica devem ser consideradas como componentes associadas ao sinal como indicado na Figura 4.6.

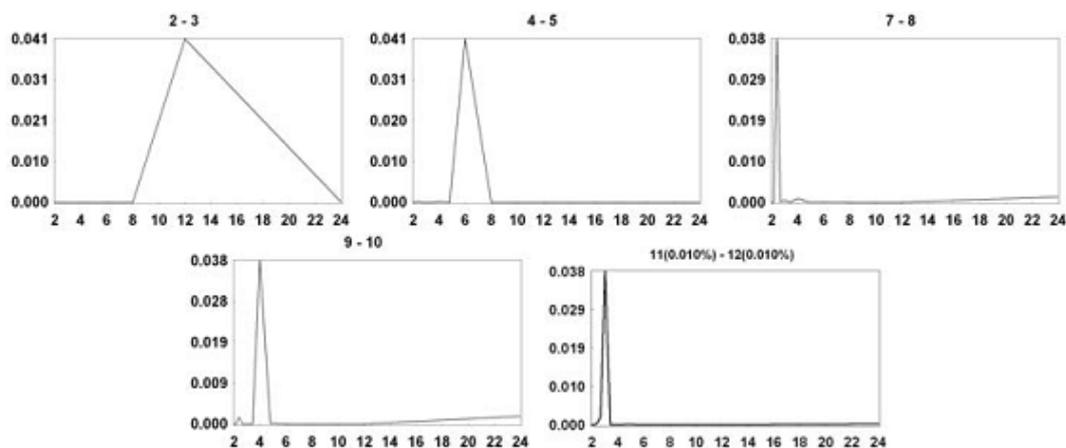


Figura 4.6 – Periodogramas dos autovalores parados (2-3, 4-5, 7-8, 9-10, 11-12).

(Fonte: Hassani, 2007)

Com base na informação complementar pode ser feito um agrupamento apropriado das autotriplas, no entanto, se o tamanho da janela L for grande, a visualização do espectro dos autovalores assim como o gráfico gerado pela matriz de correlações ponderadas será difícil. Por este motivo, no Apêndice A é apresentada a *Análise de Clusters Usando Métodos Hierárquicos Aglomerativos* como sendo uma técnica para o agrupamento das autotriplas obtidas na SVD a qual auxilia o processo de previsão da velocidade do vento.

Tendo em mente que a velocidade do vento constitui somente um dos passos da estratégia metodológica para a obtenção da previsão da densidade condicional da geração eólica, no capítulo seguinte será exposto o arcabouço teórico que se fundamenta o cálculo da estimação da densidade condicional da produção eólica, através de estimadores *kernel*,

o que em outras palavras significa o cálculo da curva de potência de forma estocástica; Além de fornecer os lineamentos teóricos para a previsão com o estimador *Nadaraya-Watson*.