

3. Singular Spectrum Analysis (SSA)

Singular Spectrum Analysis (SSA) é um poderoso método em séries temporais que incorpora elementos de análise clássica de séries temporais, estatística multivariada, geometria multivariada, sistemas dinâmicos e processamentos de sinais (GOLYANDINA et al., 2001). SSA pode ser aplicada em diversas áreas: da matemática e física à economia e matemática financeira, da meteorologia e oceanografia a ciências sociais (HASSANI, 2007).

O surgimento do SSA é geralmente associado com a publicação de trabalhos de BROOMHEAD & KING (1986). A descrição detalhada dos fundamentos teóricos e práticos da técnica SSA (com muitos exemplos) pode ser encontrada em (GOLYANDINA et al., 2001; HASSANI, 2007).

Um dos objetivos da SSA é fazer uma decomposição da série original em uma soma de um pequeno número de componentes independentes e interpretáveis como uma tendência de variação lenta, componentes oscilatórias e uma estrutura de ruído. SSA é uma ferramenta que pode ser usada para resolver os seguintes problemas: 1) encontrar tendências de diferentes resoluções; 2) suavizar séries temporais; 3) extrair componentes sazonais; 4) extrair ciclos com pequenos e grandes períodos; 5) extrair periodicidades com amplitudes variáveis; 6) extrair tendências complexas e periodicidades e 7) encontrar estrutura em séries temporais curtas (HASSANI, 2007). No caso particular desta tese, a suavização de séries temporais é tratada através da remoção de ruídos.

Uma das vantagens do método proposto em SSA é sua abordagem não paramétrica; ou seja, não é necessário conhecer o modelo paramétrico para a série temporal considerada. Uma introdução elementar ao SSA pode ser encontrada em (GOLYANDINA et al., 2001).

Por meio do método SSA, uma série temporal pode ser transformada em uma matriz trajetória (nome dado a matriz obtida com a definição do parâmetro L), que é expandida em termos da decomposição em valores singulares (HASSANI, 2007). Cada componente desta expansão concentra uma parcela da “energia” contida na matriz trajetória gerada a partir da série temporal (HAMILTON, 1994). Dessa forma,

um subconjunto de componentes concentra a maior parte da energia total com estrutura de dependência temporal e ruído; enquanto que as componentes restantes concentram a parte da energia sem qualquer estrutura de dependência temporal ou informação (isto é, são constituídas apenas de ruído). Assim sendo, com o uso de algum método de seleção de componentes, como os apresentados na seção 3.1.3, pode-se realizar a separação de tais componentes em dois grupos: um contendo as componentes que detêm a estrutura de dependência temporal e outro com as componentes que detêm apenas ruído. A soma das componentes que concentram a estrutura de dependência temporal gera uma versão aproximada e menos ruidosa da série temporal original. Isto é, por meio do método SSA e do método de seleção de componentes SSA, pode-se remover parte do ruído presente na série temporal original.

3.1. SSA

O método básico SSA consiste em dois estágios complementares: decomposição e reconstrução; ambos incluem duas etapas separadas. No primeiro estágio a série original é decomposta e no segundo estágio a série original é reconstruída menos ruidosa e usada para modelagem e previsão pontual. Um dos principais conceitos no estudo das propriedades de SSA é a 'separabilidade', o que caracteriza o quão bem diferente os componentes podem ser separados uns dos outros. A ausência de separabilidade aproximada é frequentemente observado em série com estrutura complexa. Para estas séries e para séries com estrutura especial, há diferentes maneiras de modificar SSA levando a diferentes versões, como SSA com centragem individuais e duplas, Toeplitz SSA e SSA sequencial (para mais detalhes, ver (ELSNER & TSONIS, 1996, seção 1.7).

Vale a pena notar que, apesar de alguns conceitos probabilísticos e estatísticos serem empregados nos métodos baseados em SSA, não se fazem necessárias quaisquer suposições estatísticas tais como estacionariedade da série ou normalidade dos resíduos.

3.1.1. Decomposição SSA

O estágio de decomposição pode ser subdividido em duas etapas: incorporação e decomposição em valores singulares (SVD - *Singular Value Decomposition*).

I) Incorporação:

Seja $Y_T = [y_1, \dots, y_T] \in \mathbb{R}^T$ uma série temporal com cardinalidade igual a T e $F: \mathbb{R}^T \rightarrow \mathbb{R}^{L \times K}$ um mapa invertível. Por incorporação, entende-se como sendo um procedimento no qual, dado um valor L – conhecido como comprimento de janela, uma série temporal $Y_T \in \mathbb{R}^T$ é transformada, através do mapa F , em uma matriz $X = [X_1, \dots, X_K]_{L \times K} \in \mathbb{R}^{L \times K}$, onde $X_j = [y_j, \dots, y_{j+L-1}]^T \in \mathbb{R}^L, j = 1, \dots, K$. Isto é, $Y_T \in \mathbb{R}^T \xrightarrow{F} X \in \mathbb{R}^{L \times K}$, onde $K = T - L + 1$. Nesta fase de incorporação obtém-se a matriz:

$$X = [X_1, \dots, X_K]_{L \times K} = (x_{ij})_{i,j=1}^{L,K} = \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_K \\ y_2 & y_3 & y_4 & \cdots & y_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \cdots & y_T \end{bmatrix} \quad (27)$$

A matriz X em (27) é conhecida como matriz trajetória (HASSANI, 2007) e o comprimento de janela L , que assume algum valor inteiro no intervalo $2 \leq L \leq T$ é o único parâmetro nesta fase (GOLYANDINA et al., 2001). Os vetores X_j , com $j = 1, \dots, K$ em (27) são chamados vetores L – defasados (ou simplesmente vetores defasados). Um tratamento especial será dado em relação à determinação e escolha do comprimento da janela L na seção 3.3.

A matriz X tem a característica de ser Hankel, ou seja, todos os elementos da diagonal onde $i + j = \text{constante}$ são iguais.

II) SVD:

Nesta etapa, a matriz trajetória X é decomposta em uma soma de matrizes elementares bi-ortogonais (KUBRUSLY, 2001) de posto 1. Considere o *operador normal e compacto* $S := XX^T$. Seja $\sigma(S)$ o *espectro* de S e $\{U_l U_l^T\}_{l=1}^L$ uma *resolução de identidade* sobre o *espaço de Hilbert* $(\mathbb{R}^L, \langle \cdot, \cdot \rangle)$ associada ao operador S , onde U_l é o autovetor associado ao autovalor $\lambda_l \in \sigma(S)$. Pode-se mostrar que S é um operador *semi-definido positivo*, de modo que $\lambda_l \geq 0$, para todo l . Seja $V_l := X^T U_l / \sqrt{\lambda_l}$ e considere a ordenação: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L \geq 0$ dos autovalores de S e U_1, U_2, \dots, U_L o sistema ortonormal dos autovetores de S correspondentes a estes autovalores. Note

que U_1, \dots, U_L formam uma base ortonormal do espaço coluna \mathcal{L}_d de X que é chamado *espaço trajetória*, $\mathcal{L}_d = \text{span}(U_1, \dots, U_d)$ (GOLYANDINA, 2010).

De acordo com KUBRUSLY (2001), se $S := XX^T$ e S é um operador normal, compacto e semi-definido positivo, então a matriz trajetória X pode ser expandida através da *decomposição em valores singulares* (SVD). Como esta *condição suficiente* é verdadeira, segue que a matriz trajetória X pode ser expandida via SVD, em:

$$X = \sum_{\lambda_i \in \sigma(S)} (\lambda_i)^{\frac{1}{2}} U_i V_i^T = E_1 + E_2 + \dots + E_L \quad (28)$$

Onde: $\{E_i\}_{i=1}^L := \lambda_i^{1/2} U_i V_i^T$, $\lambda_i^{1/2}$, $i = 1, \dots, L$ são os valores singulares e os conjuntos $\{\lambda_i^{1/2}\}_{i=1}^L$, $\{U_i\}_{i=1}^L$ e $\{V_i\}_{i=1}^L$ são, respectivamente, denominados *espectro singular* e *vetores singulares à esquerda e à direita* da matriz trajetória X . A coleção $(\sqrt{\lambda_i}, U_i, V_i)$ é conhecida como a *l-ésima autotripla na SVD* da matriz trajetória X . A contribuição de cada componente em (28) pode ser mensurada pela razão de valores singulares, dada por $(\lambda_i)^{1/2} / \sum_{i=1}^L (\lambda_i)^{1/2}$.

Considere d o *posto* (isto é, o número de autovalores não nulos) da matriz trajetória X . Assim:

$$d = \max\{l, \text{tal que } \lambda_l > 0\} = \text{posto}(X) \quad (29)$$

Desta forma, a identidade descrita em (28) pode ser reescrita como:

$$X = E_1 + E_2 + \dots + E_d, \quad (30)$$

onde $d \leq L$.

Segundo HASSANI (2007), outra característica da SVD está relacionada com as propriedades das direções determinadas pelo autovetores. Especificamente, o primeiro autovetor determina a direção de tal forma que a variação das projeções dos vetores defasados nessa direção é máxima. Cada autovetor subsequente determina a direção que é ortogonal a todas as direções anteriores e a variação da projeção dos vetores defasados para esta direção também é máxima. Portanto, é natural chamar a direção do l -ésimo autovetor U_l a l -ésima direção principal. Note-se que as matrizes

elementares E_l são construídas a partir das projeções dos vetores defasados nas l -ésimas direções particulares. Este ponto de vista sobre a SVD da matriz trajetória composto por vetores L - defasados e um apelo à associação com a análise de componentes principais leva a seguinte terminologia: podemos chamar o vetor U_l de o autovetor l , o vetor V_l será chamado o l -ésimo vetor fator e o vetor $Z_l = \sqrt{\lambda_l} V_l$ o l -ésimo componente principal.

3.1.2. Reconstrução SSA

O estágio de reconstrução pode ser subdividido em duas etapas: agrupamento e média diagonal. Na etapa de agrupamento, as componentes da série são separadas através de algum processo identificador destas componentes e na etapa de média diagonal, a série aproximada é reconstruída.

I) Agrupamento:

A etapa de agrupamento corresponde a dividir as matrizes elementares em vários grupos e somar as matrizes dentro de cada grupo. Este processo consiste em, dada a decomposição em (30), considerar a sequência $\{E_l\}_{l=1}^d$ de matrizes elementares nesta decomposição. Agrupe-as em $m \leq d$ grupos *disjuntos* utilizando algum método de agrupamento (por exemplo, com o auxílio da *análise de componentes principais*, *análise de agrupamentos por clusterização hierárquica integrada com análise de componentes principais* ou *análise gráfica de vetores singulares*. Estes métodos serão vistos após a definição de Média Diagonal, nas seções de 3.1.2.1 a 3.1.2.3) e assumindo que, após o agrupamento, o conjunto de índices gerado é dado por $\{I_1, \dots, I_m\}$, onde, para todo i , $I_i = \{I_{i1}, \dots, I_{ip_i}\}$ e p_i é a cardinalidade do grupo I_i . Note que $\{E_l\}_{l=1}^d = \bigcup_{i=1}^m \{E_{l_{ij}}\}_{j=1}^{p_i}$, onde $m \leq d$. A matriz elementar E_{l_i} gerada a partir do grupo $\{E_{l_{ij}}\}_{j=1}^{p_i}$ é dada por $E_{l_i} = \sum_{j=1}^{p_i} E_{l_{ij}}$, de modo que a identidade em (30) pode ser reescrita como em (31).

$$X = \sum_{i=1}^m E_{l_i} \quad (31)$$

O procedimento de agrupamento pode ser realizado também sob a sequência $\cup_{i=1}^m \{E_{l_{ij}}\}_{j=1}^{p_i}$ de igual forma ao realizado na sequência $\{E_l\}_{l=1}^d$ utilizando, inclusive, um método ou critério de agrupamento diferente do aplicado sob a sequência $\{E_l\}_{l=1}^d$ em (30). O objetivo do agrupamento é diminuir o número de componentes (ou matrizes elementares) na SVD da matriz trajetória X . A contribuição da componente E_{l_i} pode ser mensurada pela razão de valores singulares dada por $\sum_{j=1}^{p_i} (\lambda_{l_{ij}})^{1/2} / \sum_{l=1}^d (\lambda_l)^{1/2}$.

II) Média Diagonal:

O objetivo da média diagonal é transformar uma matriz para a forma de uma matriz Hankel que pode ser posteriormente convertida em uma série temporal. Se z_{ij} representa um elemento de uma matriz Z , então, o k -ésimo termo da série resultante é obtido por uma média de z_{ij} sobre todos i, j tais que $i + j = k + 1$. Este procedimento é chamado média diagonal, ou Hankelização da matriz Z . O resultado da Hankelização de uma matriz Z é a matriz Hankel $\mathcal{H}Z$.

Ao aplicar o processo de Hankelização a cada elemento E_{l_i} , $i = 1, \dots, m$, em (31), obtém-se uma nova expansão:

$$X \cong \tilde{E}_{l_1} + \tilde{E}_{l_2} + \dots + \tilde{E}_{l_m} \quad (32)$$

Onde $\tilde{E}_{l_i} = \mathcal{H}E_{l_i}$. A cada matriz desta expansão o processo de Hankelização é aplicado através da média diagonal, apresentado a seguir, e que equivale a aplicação bijetora inversa F^{-1} do mapa invertível F sobre a matriz trajetória X .

Considere a matriz trajetória X e assuma que $L^* = \min(L, K)$ e que $K^* = \max(L, K)$. Considere também que $x_{l,k}^{(i)}$ seja um elemento na linha l e coluna k na matriz E_{l_i} . O elemento $y_t^{(i)}$ da componente SSA $\left[y_t^{(i)} \right]_{1 \times T}$ é calculado por meio da *média diagonal*, definida em (33), a partir da matriz elementar E_{l_i} .

$$\tilde{y}_t^{(i)} = \begin{cases} \frac{\sum_{l=1}^t x_{l,t-l+1}^{(i)}}{t}, & \text{se } 1 \leq t < L^* \\ \frac{\sum_{l=1}^{L^*} x_{l,t-l+1}^{(i)}}{L^*}, & \text{se } L^* \leq t < K^* \\ \frac{\sum_{l=t-K^*+1}^{T-K^*+1} x_{l,t-l+1}^{(i)}}{T-K^*+1}, & \text{se } K^* \leq t \leq T \end{cases} \quad (33)$$

Este processo equivale à decomposição da série original $Y_T = [y_1, \dots, y_T]$ em uma soma de m séries.

$$y_t = \sum_{i=1}^m \tilde{y}_t^{(i)}, \quad (34)$$

onde $\tilde{E}_{I_i} = (\tilde{y}_1^{(i)}, \tilde{y}_2^{(i)}, \dots, \tilde{y}_T^{(i)})$, $i = 1, \dots, m$. A série aproximada será $\tilde{Y}_T = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_T)$, onde $y_t = \sum_{l=1}^r \tilde{y}_t^{(l)}$, $t = 1, \dots, T$ e $r < m$.

Observa-se que em termos de F , o processo de média diagonal é feito através da sequência:

$$\begin{aligned} F^{-1}(X) &= F^{-1}\left(\sum_{i=1}^m E_{I_i}\right) = F^{-1}\left(\sum_{i=1}^m \left(\sum_{j=1}^{p_i} E_{I_{ij}}\right)\right) = \sum_{i=1}^m F^{-1}\left(\sum_{j=1}^{p_i} E_{I_{ij}}\right) \Rightarrow \\ &\Rightarrow \sum_{i=1}^m \left(\sum_{j=1}^{p_i} F^{-1}(E_{I_{ij}})\right) = \sum_{i=1}^m \left(\sum_{j=1}^{p_i} [y_t^{(ij)}]_{1 \times T}\right) = \sum_{i=1}^m [y_t^{(i)}]_{1 \times T} \quad (35) \\ &= \sum_{i=1}^m [y_t^{(i)}]_{1 \times T} = [y_t]_{1 \times T}. \end{aligned}$$

Cada componente $[y_t^{(i)}]_{1 \times T}$ concentra parte da energia da série temporal original $[y_t]_{1 \times T}$ que pode ser mensurada pela razão de valores singulares $\sum_{j=1}^{p_i} (\lambda_{I_{ij}})^{1/2} / \sum_{l=1}^d (\lambda_l)^{1/2}$.

De acordo com GOLYANDINA et al. (2001), as componentes SSA $[y_t^{(i)}]_{1 \times T}$ podem ser classificadas em três categorias: *tendência*, *harmônica* e *ruído*.

3.1.3. Métodos de Agrupamentos

No estágio de reconstrução SSA, o método utilizado para se fazer o agrupamento tem uma importância fundamental. Resultados distintos podem acontecer dependendo do método usado.

Dentre os diversos métodos de agrupamento, destacam-se alguns que vem sendo utilizados em trabalhos científicos tratando este assunto e servem de base de utilização para os resultados desta tese.

3.1.3.1. Agrupamento por Análise de Componentes Principais

Considerando cada valor singular λ_l na SVD, em ordem decrescente de magnitude, que é expressa pela contribuição na SVD através da razão dos valores singulares $\sum_{j=1}^{p_i} (\lambda_{l_{ij}})^{1/2} / \sum_{l=1}^d (\lambda_l)^{1/2}$ e através da análise de componentes principais (ACP) (VASCONCELOS, 2011), escolhe-se o valor de R no conjunto de índices $\{1, \dots, m\}$ de modo que apenas as R primeiras autotriplas são consideradas na expansão (32). Assim, a série temporal aproximada filtrada SSA através de ACP é dada por:

$$[\tilde{y}_t]_{1 \times T} = \sum_{i=1}^R [y_t^{(i)}]_{1 \times T} \quad (36)$$

Neste caso, a série aproximada gerada é menos ruidosa e SSA não trabalha com remoção de componentes com tendência ou sazonalidades.

3.1.3.2. Agrupamento por Clusterização Hierárquica

Na literatura, podem ser encontrados diferentes métodos de *análise de agrupamentos* (ou *cluster analysis*) que consistem em *métodos de classificação não supervisionados* usados para encontrar uma estrutura natural de agrupamentos em objetos multidimensionais. De acordo com ALDENDERFER & BLASHFIELD (1984), a *análise de agrupamentos* visa a agrupar um conjunto com N objetos em K *clusters* mutuamente *excludentes*, de tal forma que os objetos em um mesmo *cluster* apresentem *similaridades* entre si e *dissimilaridades* em relação aos objetos pertencentes aos outros *clusters*.

Os vetores singulares resultantes na SVD podem apresentar perfis semelhantes, de modo que podem ser agrupados por meio de *análise de agrupamentos*. Qualquer um dos métodos de análise de agrupamento pode ser utilizado na classificação dos vetores singulares na SVD. Os métodos hierárquicos

agrupam um conjunto de N objetos sequencialmente em 2, 3, 4 até $N - 1$ grupos, obtendo no final uma estrutura em árvore, semelhante às classificações zoológicas - espécie, gêneros, famílias, ordem etc.

3.1.3.3. Agrupamento por Análise Gráfica dos Vetores Singulares

A análise das coordenadas da série temporal na base definida pelos vetores singulares resultantes da SVD permite identificar as componentes de tendência e da sazonalidade da série. O problema geral aqui consiste em identificar e separar as componentes oscilatórias das componentes que fazem parte da tendência. De acordo com GOLYANDINA et al. (2001) a análise gráfica de tais coordenadas aos pares permite identificar por meio visual as componentes harmônicas da série.

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 comprimento da janela $Le K$. Se o parâmetro ρ assume um valor inteiro, então ρ é classificado como *período do harmônico* (MORETTIN & TOLOI, 2006). As coordenadas da série temporal em duas componentes ortogonais podem ser dispostas em um diagrama de dispersão.

As funções seno e o cosseno com frequências, amplitudes e fases iguais resultam em um diagrama de dispersão que exibe um padrão circular (MORETTIN & TOLOI, 2006). Por sua vez, se $\rho = \frac{1}{\omega}$ é um inteiro, então o diagrama de dispersão exibe um polígono regular com ρ vértices. Para uma frequência $\omega = m/n < 0.5$ com m e n inteiros e primos entre si, os pontos são vértices de um polígono regular de n vértices (GOLYANDINA et al., 2001). Dessa forma, a identificação dos componentes que são gerados por um harmônico é reduzida à análise pictórica do padrão determinado nos diferentes pares de componentes.

Em CASSIANO et al. (2013) há uma aplicação da filtragem SSA aplicada à série de vazão que compara os métodos de agrupamento citados e os resultados mostram que o poder preditivo melhora quando se utiliza a análise gráfica dos vetores singulares. Desta forma, esta metodologia será utilizada na fase de agrupamento SSA/MSSA nesta tese.

3.1.3.4. Informação Suplementar: Análise de Periodograma

A análise de periodograma auxilia na identificação de uma componente harmônica geral. De acordo com MIRANIAN et al. (2013), o periodograma do vetor singular de cada autotripla fornece informações sobre o comportamento periódico do componente e frequência (período) das oscilações. Portanto, o agrupamento adequado pode ser feita com a ajuda de análise de periodograma.

Para a série $Y_T = (y_1, \dots, y_T)$, o periodograma $\Pi_{Y_T}^T(\omega)$ é definido por:

$$\Pi_{Y_T}^T(\omega) = \frac{1}{T} \left| \sum_{t=0}^{T-1} e^{-2\pi i \omega t} y_{t+1} \right|, \quad \omega = (-0.5, 0.5] \quad (37)$$

onde $i = \sqrt{-1}$.

Para as componentes periódicas, o periodograma tem pontas afiadas em torno de frequência do componente (período). Por isso, a identificação visual é simples.

3.1.4. Separabilidade

Por correlação ponderada (*w-correlation*), entende-se a função que quantifica a dependência linear entre duas componentes SSA $Y_t^{(i)}$ e $Y_t^{(j)}$, conforme definido por

$$\rho_{ij}^{(w)} = \frac{(Y_t^{(i)}, Y_t^{(j)})_w}{\|Y_t^{(i)}\|_w \|Y_t^{(j)}\|_w}, \quad (38)$$

onde $\|\cdot\|_w$ é a norma euclidiana, $(\cdot)_w$ é o produto interno tal que: $\|Y_t^{(i)}\|_w = \sqrt{(Y_t^{(i)}, Y_t^{(i)})_w}$ e $(Y_t^{(i)}, Y_t^{(j)})_w = \sum_{k=1}^T w_k y_k^{(i)} y_k^{(j)}$; e $w_k = \min\{k, L, T - k\}$.

Por meio da separabilidade, pode-se verificar estatisticamente se as duas componentes SSA estão bem separadas, em termos de dependência linear. Se o valor absoluto w é pequeno (HASSANI, 2007), então as componentes SSA correspondentes são classificadas como *w*-ortogonais (ou quase *w*-ortogonais); caso contrário, são ditas mal separadas. Salienta-se que comumente utiliza-se a correlação

ponderada na fase de agrupamento para geração de novas matrizes elementares na SVD (GOLYANDINA et al., 2001). A matriz contendo os valores absolutos das correlações w correspondentes para a decomposição completa pode fornecer informação útil para o agrupamento correto das autotriplas (MIRANIAN et al., 2013).

3.2. Multi-channel Singular Spectrum Analysis (MSSA)

SSA é um método em expansão e seu uso é aplicado a diversas áreas do conhecimento. O problema de filtragem de sinais univariados menos ruidosos pode ser tratado perfeitamente com o uso de SSA, como em HASSANI (2007), por exemplo. Para problemas de reconstrução e previsão de sinais em um sistema de séries temporais simultâneas, uma extensão do SSA pode ser usada para a solução: o *Multi-channel SSA* (MSSA).

MSSA é uma extensão do SSA para trabalhar com análise e previsão de séries temporais multidimensionais (GOLYANDINA & STEPANOV, 2005).

O procedimento MSSA segue a mesma estrutura do procedimento SSA com a diferença de fazê-lo usando um sistema de séries temporais em face de uma única série usada no SSA.

Considere o sistema de s séries temporais de tamanho T

$$Y^{(k)} = \left(y_t^{(k)} \right)_{t=1}^T \quad (39)$$

Onde $k = 1, \dots, s$. O caso particular do procedimento MSSA para $s = 1$, equivale ao procedimento usando SSA (GOLYANDINA & STEPANOV, 2005).

Escolhendo um único comprimento de janela de defasagem L para todas as séries, onde $1 < L < T$, na fase de incorporação MSSA são obtidos $K = T - L + 1$ vetores defasados $X_j^{(k)} = \left(y_j^{(k)}, \dots, y_{j+L-1}^{(k)} \right)'$, $j = 1, \dots, K$ para cada série $Y^{(k)}$, $k = 1, \dots, s$. Assim, para cada série $Y^{(k)}$, é possível obter através de um mapa invertível $F^{(k)}$, uma matriz trajetória como em (39).

$$X^{(k)} = \begin{bmatrix} y_1^{(k)} & y_2^{(k)} & \cdots & y_K^{(k)} \\ y_2^{(k)} & y_3^{(k)} & \cdots & y_{K+1}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ y_L^{(k)} & y_{L+1}^{(k)} & \cdots & y_T^{(k)} \end{bmatrix} \quad (40)$$

De acordo com HASSANI & MAHMOUDVAND (2013), a matriz trajetória da série multidimensional $(Y^{(1)}, Y^{(2)}, \dots, Y^{(s)})$ pode ser descrita de forma vertical VMSSA ou horizontal HMSSA. Quando o tratamento é vertical, K permanece fixo, mas L pode ser variável. Neste caso, uma matriz de dimensão $Ls \times K$ é obtida como em (41):

$$\tilde{X} = [X_1^{(1)} : \dots : X_K^{(1)} : \dots : X_1^{(s)} : \dots : X_K^{(s)}]' = [X^{(1)} : \dots : X^{(s)}]' \quad (41)$$

Assim:

$$\tilde{X} = \begin{bmatrix} X^{(1)} \\ X^{(2)} \\ \vdots \\ X^{(s)} \end{bmatrix} = \begin{bmatrix} X_1^{(1)} & X_2^{(1)} & \cdots & X_K^{(1)} \\ X_1^{(2)} & X_2^{(2)} & \cdots & X_K^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ X_1^{(s)} & X_2^{(s)} & \cdots & X_K^{(s)} \end{bmatrix}. \quad (42)$$

Depois disso, a matriz ampliada de covariância C_X de dimensões $Ls \times Ls$ é calculada como em (43):

$$C_X = \frac{1}{K} \tilde{X} \tilde{X}' = \begin{bmatrix} C_1^{(1)} & C_2^{(1)} & \cdots & C_s^{(1)} \\ C_1^{(2)} & C_2^{(2)} & \cdots & C_s^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ C_1^{(s)} & C_2^{(s)} & \cdots & C_s^{(s)} \end{bmatrix}, \quad (43)$$

onde $C_k^{(k')} = \frac{1}{K} X^{(k)} (X^{(k')})'$ para todo $k, k' = 1, 2, \dots, s$. Uma vez que $X^{(k)}$ é uma matriz $L \times K$, então $C_k^{(k')}$ é uma matriz $L \times L$. Também é possível notar que C_X é uma matriz simétrica. Cada bloco $C_k^{(k')}$ contém uma estimativa da covariância entre os dois canais k e k' . Ao diagonalizar a matriz C_X de dimensão $Ls \times Ls$, obtém-se Ls autovalores e seus respectivos autovetores.

Ainda segundo HASSANI & MAHMOUDVAND (2013), quando a abordagem HMSSA, então a matriz \tilde{X} terá dimensão $L \times Ks$ e, neste caso, L permanece fixo e K pode ser variável. A matriz trajetória tem formato:

$$\tilde{X} = [X_1^{(1)} : \dots : X_K^{(1)} : \dots : X_1^{(s)} : \dots : X_K^{(s)}] = [X^{(1)} : \dots : X^{(s)}] \quad (44)$$

O *espaço trajetória* (GOLYANDINA, 2010) é definido por um espaço linear gerado pelos vetores defasados (colunas da matriz trajetória \tilde{X}).

Como \tilde{X} tem dimensão $L \times Ks$, então $S = \tilde{X}\tilde{X}'$ como definido em KUBRUSLY (2001) tem dimensão $L \times L$ e então pode-se obter o SVD da matriz \tilde{X} de forma análoga ao SSA.

De forma análoga, a fase de agrupamento particiona o conjunto de índices $\{1, \dots, d\}$ em m subconjuntos disjuntos I_1, \dots, I_m de modo que a matriz trajetória \tilde{X} pode ser reescrita no que pode ser chamado de decomposição agrupada:

$$\tilde{X} = \tilde{X}_{I_1} + \dots + \tilde{X}_{I_m} \quad (45)$$

Por fim a média diagonal é aplicada a cada uma das séries decomposta em (45) e então o grupo de sinais reconstruídos, dado por $\tilde{Y}^{(k)} = \left(\tilde{Y}_t^{(k)}\right)_{t=1}^T$, $k = 1, \dots, s$ é obtido.

Apesar de haver duas abordagens em MSSA com a possibilidade de variação em L ou em K , nas aplicações desta tese são consideradas séries de mesmo comprimento (T), conseqüentemente, os parâmetros L e K devem permanecer fixos utilizando a abordagem HMSSA.

3.3. Escolha do valor do parâmetro L de defasagem

Desde o início dos estudos em SSA, a determinação de um valor para o parâmetro L de defasagem vem sendo alvo de discussão em diversos artigos por diversos autores. Em GOLYANDINA (2010) há uma longa discussão acerca do comprimento ideal da janela de defasagem partindo do princípio que este valor pode ser fixado ou variável, dependendo de seu comportamento assintótico quando $T \rightarrow \infty$.

Ainda observando o comportamento assintótico do comprimento da janela (*window length*) a considerar a matriz $L \times K$, com $K \rightarrow \infty$, a escolha de um número fixo de linhas L não muito grande tem um custo operacional menor (GOLYANDINA, 2010), muito embora hajam recursos computacionais avançados recentes como os mencionados em KOROBEYNIKOV (2010), que podem fazer cálculos muito rápidos. Em GOLYANDINA (2010) verificou-se que no processo SSA, no qual se pode separar bem o sinal S_T do ruído R_T , se o sinal for do tipo

$$S_T = a^T \cos\left(\frac{\pi T}{5}\right), \quad (46)$$

a escolha de um nível de ruído é tal que se tem alguma razão sinal-ruído (SNR – *signal-to-noise ratio*) dada por: $SNR = \frac{\sum_{t=1}^T S_t^2}{\sum_{t=1}^T R_t^2}$ ou $SNR = \frac{\sum_{t=1}^T S_t^2}{\sigma_{R_T}^2}$ se os resíduos são aleatórios. SNR pode ser usado para comparar a qualidade do processamento de séries de igual tamanho. Entretanto, não podemos dizer que SSA separa sinal de ruído apenas se SNR for maior que um determinado valor, por exemplo, para qualquer SNR pequeno, um sinal de onda senoidal é assintoticamente separável de um ruído branco com $T \rightarrow \infty$ e $L \sim \beta T$ com $0 < \beta < 1$. Este mesmo artigo mostra, através de experimentos, que para séries aplicadas com sinal como em (40), a perturbação do sinal cresce quando L cresce e decresce quando $K = T - L + 1$ cresce. Então os erros resultantes são causados por estas tendências contraditórias. Mostra também que o crescimento da razão dos erros é maior para $L > N/2$. No entanto, não é surpresa que os erros reconstruídos são grandes para comprimento de janela L pequeno. Com isso, o artigo mantém a premissa que a questão sobre o comprimento da janela permanece em aberto.

Em várias aplicações, a dependência dos erros reconstruídos é levada em consideração para determinação do comprimento da janela. Em GOLYANDINA & VLASSIEVA (2009) há uma consideração sobre o caso de um sinal constante ruidoso $S_T \equiv c$. Neste artigo é explicitada a variância dos erros reconstruídos. Os resultados lá apresentados levaram em consideração a forma assintótica da variância dos erros de primeira ordem apresentada pela fórmula da dependência assintótica dos erros ao comprimento da janela.

Considere o comprimento da janela $L = \beta T$, $0 < \beta \leq 1/2$, e l um índice pontual da série tal que $l \sim \frac{\gamma T}{2}$, $0 \leq \gamma \leq 1$, com $T \rightarrow \infty$. Quando $\gamma = 1$, o parâmetro está no meio da série, conseqüentemente a fórmula para a primeira metade da série com comprimentos de janela L menor que a metade do comprimento da série é apresentada. Daí, a variância dos erros de primeira ordem terá a seguinte forma assintótica (GOLYANDINA, 2010):

$$\mathbb{D}_{S_l}^{(1)} \sim \frac{\sigma^2}{T} \begin{cases} D_1(\beta, \gamma), & 0 \leq 2 \min(\beta, 1 - 2\beta), \\ D_2(\beta, \gamma), & 2 \min(\beta, 1 - 2\beta) < \gamma < 2\beta, \\ D_3(\beta, \gamma), & 2\beta \leq \gamma \leq 1. \end{cases} \quad (47)$$

com $T \rightarrow \infty$, onde

$$D_1(\beta, \gamma) = \frac{1}{12\beta^2(1-\beta)^2} (\gamma^2(1+\beta) - 2\gamma\beta(1+\beta)^2 + 4\beta(3-3\beta+2\beta^2)),$$

$$D_2(\beta, \gamma) = \frac{1}{6\beta^2(1-\beta)^2\gamma^2} (\gamma^4 + 2\gamma^3(3\beta-2-3\beta^2) + 2\gamma^2(3-9\beta+12\beta^2-4\beta^3) + 4\gamma(-1+4\beta-3\beta^2-4\beta^3+4\beta^4) + (8\beta-56\beta^2+144\beta^3-160\beta^4+64\beta^5)),$$

$$D_3(\beta, \gamma) = \frac{2}{3\beta}.$$

Os pontos de escolhas entre os casos em (47) correspondem a $l = K - L$, ou seja: $\gamma = 2(1 - 2\beta)$ e $l = L(\gamma - 2\beta)$. O primeiro ponto da escolha está presente se $K < 2L$ ($\beta > 1/3$). Note que a fórmula (47) pode ser estendida para o comprimento de janela $2 < L < T - 1$ ($0 < \beta < 1$) e os índices da série pontual $0 \leq l \leq T - 1$ ($0 \leq \gamma \leq 2$) dada a simetria dos erros com respeito ao ponto médio da série temporal e a equivalência dos resultados sob a substituição de L por $K(\beta \leftrightarrow 1 - \beta)$ (GOLYANDINA, 2010).

Segundo GOLYANDINA & VLASSIEVA (2009), ao tentar resolver o problema de minimizar a raiz quadrada da soma dos quadrados dos erros (*RMSE – Root Mean Square Error*) do estimador de S_l em um ponto fixo l , o comprimento da janela varia entre $T/3$ e $T/2$. Isto significa que sempre que se está com o caso com o sinal constante, o comprimento da janela, que minimiza os erros de reconstrução da série como um todo, depende do peso de cada ponto na série temporal. Em todo caso, a recomendação é a de se escolher o comprimento da janela um pouco menor que a metade do comprimento da série T . Assim, escolher o comprimento da janela usando $T/3 \leq L \leq T/2$ faz sentido.

Mesmo com todas as especulações e programas que tentem simular um valor para o comprimento da janela, os autores indicam que a escolha deste parâmetro sempre vai depender da natureza dos dados. Portanto, sempre haverá um intervalo de confiança em que o valor de L deverá se adequar.

Pensando nesta necessidade de se obter este valor de L e sabendo destas premissas anteriores sobre o tema, este trabalho propõe uma metodologia para se obter esta estimativa baseando-se nas estatísticas do teste *BDS* que foi estabelecido

em BROCK et al. (1987) e testa a independência linear e não linear entre os dados da série baseando-se na dimensão da correlação entre os mesmos. Na ocasião, valores decrescentes para L (a partir de $(T + 1)/2$ se T for ímpar ou a partir de $T/2$ se T for par) são testados até se encontrar um valor L tal que, ao se testar via *BDS* a séries de resíduos obtida na fase de agrupamento *SSA*, o teste não rejeite a hipótese de independência dos dados nas dimensões medidas nele ao nível de 5% de significância. Este critério é ilustrado nas aplicações a energia eólica e ENA desta tese.