

MODELAGEM AUTOREGRESSIVA PERIÓDICA DE SÉRIES TEMPORAIS DE VELOCIDADE DO VENTO COM ABORDAGEM MSSA

Moisés Lima de Menezes

Departamento de Estatística - Universidade Federal Fluminense
Rua Mário Santos Braga, S/N, Campus Valonguinho, Centro, Niterói - RJ, 24.220-900.
moises_lima@msn.com

Reinaldo Castro Souza

Departamento de Engenharia Elétrica - Pontifícia Universidade Católica do Rio de Janeiro
Rua Marquês de São Vicente, 225, Gávea, Rio de Janeiro – RJ, 22.451-900.
reinaldo@ele.puc-rio.br

José Francisco Moreira Pessanha

Instituto de Matemática e Estatística - Universidade do Estado do Rio de Janeiro
Rua São Francisco Xavier, 524, Maracanã, Rio de Janeiro – RJ, 20.550-013.
professorjfm@hotmai.com

Resumo

A análise espectral singular (SSA - *Singular Spectrum analysis*) é uma técnica que decompõe uma série temporal em um conjunto de componentes tais como tendência, harmônicos e ruídos. Deixando de fora as componentes ruidosas e somando as outras, a série temporal pode ser suavizada. O modelo autorregressivo periódico PAR(p) vem sendo amplamente utilizado em modelagem de séries de Energia Natural Afluente. Este artigo apresenta uma abordagem deste método de decomposição, na qual a modelagem PAR (p) é aplicada na série suavizada obtida pela SSA e por sua versão multivariada, denominada MSSA. Para ilustrar a aplicação da metodologia proposta, foram consideradas séries de velocidade do vento registradas em duas localidades na região Nordeste do Brasil. Os resultados obtidos mostram que a modelagem PAR(p) é beneficiada pela suavização prévia da série temporal por SSA/MSSA.

Palavras Chave: MSSA, modelagem PAR(p), velocidade do vento.

Abstract

Singular spectrum analysis (SSA) is a technique that decomposes a time series into a set of components, such as, trend, harmonics, and residuals. Leaving out the residual components and adding up the others, the time series can be smoothed. The PAR(p), periodic autoregressive models, has been broadly used in modeling Affluent natural Energy series. This article presents an approach of this decomposition method, in which the PAR (p) model is applied to the smoothed series obtained by SSA and its multivariate version, named MSSA. To illustrate the application of the proposed methodology were considered wind speed time series recorded at two locations in northeastern Brazil. The results show that the PAR (p) model is favored by prior smoothing of the time series by SSA / MSSA.

Keywords: MSSA, PAR(p) modeling, wind speed.

1. INTRODUÇÃO

Com o advento da energia eólica, a operação dos sistemas elétricos deverá contar com modelos capazes de prever a produção de energia elétrica nos parques eólicos, a fim de proporcionar uma integração segura e econômica dos parques na operação do Sistema Interligado Nacional (SIN). Este trabalho descreve uma aplicação do procedimento Singular Spectrum Analysis (SSA), combinado com o modelo periódico autorregressivo PAR (p), na modelagem da série temporal da média mensal da velocidade do vento.

SSA é um poderoso método para análise e previsão de séries temporais e incorpora elementos de análise de séries temporais clássica, estatística multivariada, geometria multivariada, sistemas dinâmicos e processamento de sinais (GOLYANDINA et al., 2001). SSA podem ser aplicados em vários campos da matemática e da física à economia e matemática financeira, meteorologia e oceanografia para ciências sociais (HASSANI, 2007).

Com base na decomposição em valores singulares (*singular value decomposition* – SVD), o método SSA decompõe uma série de tempo em uma soma de componentes, tal que em cada componente se concentra uma parte da energia contida na série temporal. Além disso, um pequeno conjunto de componentes ocorre na maior parte da estrutura de autocorrelação contida na série temporal, enquanto que os restantes podem ser considerados como componentes de ruído. Portanto, o conjunto de componentes pode ser classificado em dois grupos mutuamente exclusivos e exaustivos: sinal e ruído. O primeiro grupo contém os componentes que possuem algumas informações sobre a estrutura da série, enquanto o último abrange os componentes de ruído. Assim, uma versão suavizada da série temporal pode ser obtida pela soma das componentes de sinal. Tradicionalmente, uma série de tempo pode ser expressa como a soma de componentes: tendência, harmônica e ruído. A remoção do componente de ruído pode contribuir para uma melhor identificação do processo estocástico subjacente da série temporal e, conseqüentemente, melhorar a precisão das previsões. JAYAWARDENA & GURUNG (2000) mostram que a remoção de ruídos em série hidrológica melhora a precisão das previsões. Uma boa revisão do assunto é apresentada em (ELSHORBAGY et al., 2002).

Multi-channel Singular Spectrum Analysis (MSSA) é a extensão natural do SSA para uma série temporal multivariada. Os procedimentos de SSA e MSSA oferecem um método de pré-tratamento de dados. Portanto, ambos os processos podem ser utilizados a fim de melhorar o ajuste de modelos de previsão e ajudar a melhorar a precisão das previsões (HASSANI & ZHIGLJAVSKY, 2009). Mas quando o interesse reside na previsão de um conjunto de séries temporais múltiplas qual procedimento é o mais adequado? Temos de aplicar o procedimento de SSA para cada série temporal individualmente ou aplicamos o procedimento MSSA ao conjunto de séries temporais? Para responder as perguntas acima, neste artigo são apresentados os resultados de um experimento computacional onde os procedimentos de SSA e MSSA foram aplicados a séries temporais de velocidade média mensal do vento em Petrolina e Pesqueira, ambos localizados na região do Nordeste.

Em seguida, o modelo auto-regressivo periódico PAR (p), foi adaptado para cada série temporal suavizada obtida por SSA e MSSA. Os resultados mostram que a SSA contribui para o ajuste de PAR(p) no modelo de cada série velocidade do vento, porém a contribuição da MSSA é ainda maior. A opção pelo modelo PAR(p) deve-se ao seu uso nos modelos computacionais que apóiam o Planejamento da Operação de Médio e Curto prazo do SIN (MACEIRA et al., 2006).

O presente artigo está organizado em seis seções. A seguir, na seção 2, há uma breve descrição da SSA. Na seção 3 tem-se uma descrição resumida da MSSA. A modelagem PAR(p) é introduzida na seção 4. Os resultados do experimento computacional são descritos na seção 5, em particular são comparados os erros de previsão alcançados pela modelagem PAR(p) das séries de velocidade do vento antes e após a filtragem por MSSA e SSA. Por fim, na seção 6 são resumidas as principais conclusões do trabalho.

2. SINGULAR SPECTRUM ANALYSIS

A versão básica do método SSA pode ser dividida em duas etapas: *decomposição* e *reconstrução*.

A etapa da *decomposição* pode ser subdividida em *incorporação* e *decomposição em valores singulares* (SVD - *Singular Value Decomposition*).

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 uma série temporal $Y_T \in \mathbb{R}^T$ é transformada pelo mapa F em uma matriz $X = [X_1, \dots, X_T]_{L \times K} \in \mathbb{R}^{L \times K}$, onde $X_k = [y_k, \dots, y_{k+L-1}]^T \in \mathbb{R}^L$, para todo k . Isto é, $Y_T \in \mathbb{R}^T \xrightarrow{F} X \in \mathbb{R}^{L \times K}$, onde $K = T - L + 1$. A matriz X é conhecida como *matriz trajetória* (HASSANI, 2007) e o parâmetro L , que assume algum valor inteiro no intervalo $2 \leq L \leq T$ é o *tamanho da janela* da matriz trajetória (GOLYANDINA et al., 2001). A matriz trajetória X pode ser expandida via *decomposição em valores singulares*, em (1).

$$X = \sum_{\lambda_l \in \sigma(S)} (\lambda_l)^{\frac{1}{2}} U_l V_l^T = \sum_{l=1}^L E_l \quad (1)$$

Onde: $E_l := \lambda_l^{1/2} U_l V_l^T$ e os conjuntos $\{\lambda_l^{1/2}\}_{l=1}^L$ e $\{U_l\}_{l=1}^L$ são, respectivamente, denominados por *espectro singular* e de *vetores singulares* da matriz trajetória X . A coleção (λ_l, U_l, V_l) é conhecida como *autotripla na SVD* da matriz trajetória X . A contribuição de cada componente em (1) pode ser mensurada pela razão de valores singulares, dada por $(\lambda_l)^{1/2} / \sum_{l=1}^L (\lambda_l)^{1/2}$. Considere que d seja o *posto* (isto é, o número de autovalores não nulos) da matriz trajetória X . Segue que a identidade descrita em (1) pode ser reescrita tal como:

$$X = \sum_{l=1}^d E_l, \text{ onde } d \leq L. \quad (2)$$

A etapa de *reconstrução* pode ser subdividida em *agrupamento* e *média diagonal*. A etapa de *agrupamento* consiste no procedimento de agrupar algumas sequências de matrizes elementares resultantes da decomposição SVD em grupos *disjuntos* e, após isso, somá-las, gerando novas matrizes elementares.

Considerando a sequência $\{E_l\}_{l=1}^d$ de matrizes elementares na SVD, em (2). Agrupando-as em $m \leq d$ grupos *disjuntos* 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 , então a matriz elementar X_{I_i} gerada a partir do grupo $\{X_{I_{ij}}\}_{j=1}^{p_i}$ é dada por $X_{I_i} = \sum_{j=1}^{p_i} X_{I_{ij}}$, de modo que a identidade em (2) pode ser reescrita como em (3).

$$X = \sum_{i=1}^m X_{I_i} \quad (3)$$

É importante salientar que o procedimento de agrupamento pode ser realizado também sob a sequência $\bigcup_{i=1}^m \{X_{I_{ij}}\}_{j=1}^{p_i}$ 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 (2). Um dos objetivos da análise de agrupamentos é separar as componentes do sinal das componentes do ruído. A contribuição da componente X_{I_i} pode ser mensurada pela razão de valores singulares dada por $\sum_{j=1}^{p_i} (\lambda_{I_{ij}})^{1/2} / \sum_{l=1}^d (\lambda_l)^{1/2}$. Tome o conjunto $\bigcup_{i=1}^m \{X_{I_{ij}}\}_{j=1}^{p_i}$ de matrizes elementares na SVD. Dado que F é um mapa invertível, segue que F^{-1} sobre a matriz trajetória X é tal que:

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

A componente $\left[y_t^{(i)}\right]_{1 \times T}$ é conhecida como *componente SSA* da série temporal $[y_t]_{1 \times T}$ gerada, por meio do mapa F^{-1} , a partir da matriz elementar X_{I_i} .

Dada a matriz trajetória X e assumindo que $L^* = \min(L, K)$ e que $K^* = \max(L, K)$. Considere que $x_{l,k}^{(i)}$ seja um elemento na linha l e coluna k na matriz X_{I_i} . O elemento $y_t^{(i)}$ da componente SSA $\left[y_t^{(i)}\right]_{1 \times T}$ é calculado por meio da *média diagonal*, que é definida em (4), a partir da matriz elementar X_{I_i} .

$$y_t^{(i)} = \begin{cases} \frac{\sum_{l=1}^t x_{l,t-l+1}^{(i)}}{t}, & \text{se } 1 \leq t < L^* \\ \frac{\sum_{l=1}^{L^*} x_{l,t-l+1}^{(i)}}{L^*}, & \text{se } L^* \leq t < K^* \\ \frac{\sum_{l=t-K^*+1}^{T-K^*+1} x_{l,t-l+1}^{(i)}}{T-K^*+1}, & \text{se } K^* \leq t \leq T \end{cases} \quad (4)$$

Cada componente $\left[y_t^{(i)}\right]_{1 \times T}$ 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 $\left[y_t^{(i)}\right]_{1 \times T}$ podem ser classificadas em três categorias: *tendência*, *componentes harmônicas* e *ruído*.

Por correlação ponderada (w), entende-se como sendo a função que quantifica a dependência linear entre duas componentes SSA $Y_t^{(i)}$ e $Y_t^{(j)}$, conforme definido por $\rho_{ij}^{(w)} = \frac{(Y_t^{(i)}, Y_t^{(j)})_w}{\|Y_t^{(i)}\|_w \|Y_t^{(j)}\|_w}$, onde $\| \cdot \|$ é a norma euclidiana, $(\cdot)_w$ é o produto interno tal que:

$$\|Y_t^{(i)}\|_w = \sqrt{(Y_t^{(i)}, Y_t^{(i)})_w} \text{ e } (Y_t^{(i)}, Y_t^{(j)})_w = \sum_{k=1}^T w_k y_k^{(i)} y_k^{(j)}; \text{ e } w_k = \min\{k, L, T - k\}.$$

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 $\rho_{ij}^{(w)}$ é pequeno (HASSANI, 2007), então as componentes SSA correspondentes são classificadas como w -ortogonais (ou quase w -ortogonais); caso contrário, são ditas mal separadas.

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.

As coordenadas da série temporal em duas componentes ortogonais podem ser dispostas em um diagrama de dispersão. 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 . Se o parâmetro ρ assume um valor inteiro, então ρ é classificado como *período do harmônico*. Por exemplo, as funções seno e o cosseno com frequências, amplitudes e fases iguais resultam em um diagrama de dispersão que exibe um padrão circular. 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, 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.

3. MULTI-CHANNEL SINGULAR SPECTRUM ANALYSIS

MSSA é uma extensão do SSA para trabalhar com análise e previsão de séries temporais multidimensionais. O procedimento MSSA segue a mesma estrutura do procedimento SSA com a diferença de fazê-lo usando um conjunto de séries temporais em face de uma única série

Considere o sistema de s séries temporais de tamanho T

$$Y^{(k)} = \left(y_t^{(k)} \right)_{t=1}^T \quad (5)$$

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

Escolhendo um único tamanho de janela de defasagem L para todas as s séries, onde $1 < L < T$, na fase de incorporação MSSA são obtidos $K = T - L + 1$ vetores defasados $X_j^{(k)} = \left(y_j^{(k)}, \dots, y_{j+L-1}^{(k)} \right)^T$, $j = 1, \dots, K$ 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 (6).

$$X^{(k)} = \begin{bmatrix} y_1^{(k)} & y_2^{(k)} & \dots & y_K^{(k)} \\ y_2^{(k)} & y_3^{(k)} & \dots & y_{K+1}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ y_L^{(k)} & y_{L+1}^{(k)} & \dots & y_T^{(k)} \end{bmatrix} \quad (6)$$

A matriz trajetória da série multidimensional $(Y^{(1)}, Y^{(2)}, \dots, Y^{(s)})$ é, então, uma matriz de dimensão $Ls \times K$ e tem a seguinte forma:

$$X = \left[X_1^{(1)} : \dots : X_K^{(1)} : \dots : X_1^{(s)} : \dots : X_K^{(s)} \right]^T = \left[X^{(1)} : \dots : X^{(s)} \right]^T \quad (7)$$

O *espaço trajetória* (GOLYANDINA et al. (2010)) é definido por um espaço linear *spanado* pelos vetores defasados (colunas da matriz trajetória X).

A partir de $S = XX^T$, onde os autovalores de S em ordem de magnitude $\lambda_1 \geq \dots \geq \lambda_{Ls} \geq 0$, U_1, \dots, U_{Ls} os respectivos autovalores associados, $d = \max \{j: \lambda_j > 0\}$ o posto da matriz S e $V_j = \frac{X^T U_j}{\sqrt{\lambda_j}}$, $j = 1, \dots, d$. Denotando $E_j = \sqrt{\lambda_j} U_j V_j^T$, então o SVD da matriz trajetória X pode ser escrito por: $X = E_1 + \dots + E_d$.

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 seja reescrita na forma conhecida como decomposição agrupada:

$$X = X_{I_1} + \dots + X_{I_m} \quad (8)$$

Por fim a média diagonal é aplicada a cada uma das séries decomposta em (8) 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.

4. MODELOS AUTORREGRESSIVOS PERIÓDICOS

Segundo HIPEL & McLEOD (1994), algumas séries temporais têm uma estrutura de autocorrelação que depende não somente do intervalo de tempo entre as observações, mas também do período observado. Estas séries podem ser analisadas por formulações autoregressivas cujos parâmetros têm comportamento periódico. Estes modelos são chamados de auto-regressivos periódicos PAR(p) onde p é um vetor apresentado por $p = (p_1, p_2, \dots, p_s)$, s é o período considerado: se for mensal, $s = 12$, se for trimestral, $s = 4$. Para os estudos desta proposta serão considerados períodos mensais. O modelo PAR(p) pode ser representado através da padronização das observações no modelo AR(p) como em (9).

$$\left(\frac{Y_t - \mu_m}{\sigma_m} \right) = \varphi_1^m \left(\frac{Y_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \varphi_2^m \left(\frac{Y_{t-2} - \mu_{m-2}}{\sigma_{m-2}} \right) + \dots + \varphi_{p_m}^m \left(\frac{Y_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + a_t^m \quad (9)$$

onde Y_t é a série sazonal de período $s = 12$, $t = 1, \dots, T$, $m = 1, \dots, s$, μ_m é a média sazonal de período s , p_m é a ordem do operador autorregressivo de período m – neste caso, a ordem muda de acordo com o período, e a_t^m é a série de resíduos independentes e identicamente distribuídos com média zero e variância $\sigma_a^{2(m)}$.

A ideia consiste no ajuste de um modelo autorregressivo de ordem p_m para cada um dos meses da série original. A primeira etapa consiste na identificação das ordens p_m mais apropriadas aos operadores autorregressivos de cada período. Esta etapa pode ser realizada a partir das funções de autocorrelação e de autocorrelação parcial.

A estrutura de dependência temporal da série pode ser descrita pelo conjunto de funções de autocorrelação ρ_k^m dos períodos $m = 1, 2, \dots, s$:

$$\begin{aligned} \rho_k^m &= E \left[\left(\frac{Y_t - \mu_m}{\sigma_m} \right) \left(\frac{Y_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] \\ &= \varphi_1^m E \left[\left(\frac{Y_t - \mu_{m-1}}{\sigma_{m-1}} \right) \left(\frac{Y_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] + \dots \\ &+ \varphi_{p_m}^m E \left[\left(\frac{Y_t - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) \left(\frac{Y_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] \\ &+ E \left[a_t^m \left(\frac{Y_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] \end{aligned} \quad (10)$$

Fixando m e variando k de 1 a p_m em (10) obtém-se para cada período um conjunto de equações periódicas de Yule-Waker.

$$\begin{bmatrix} 1 & \rho_1^{m-1} & \dots & \rho_{k-1}^{m-1} \\ \rho_1^{m-1} & 1 & \dots & \rho_{k-2}^{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1}^{m-1} & \rho_{k-2}^{m-2} & \dots & 1 \end{bmatrix} \begin{bmatrix} \varphi_{k1}^m \\ \varphi_{k2}^m \\ \vdots \\ \varphi_{kk}^m \end{bmatrix} = \begin{bmatrix} \rho_{k1}^m \\ \rho_{k2}^m \\ \vdots \\ \rho_{kk}^m \end{bmatrix} \quad (11)$$

O conjunto de valores φ_{kk}^m , $m = 1, \dots, s$ é denominado autocorrelação parcial do período m . Cada coeficiente de autocorrelação parcial de ordem k coincide com o último parâmetro de um modelo autorregressivo da mesma ordem. Assim, em um processo autorregressivo de ordem p_m , a função de autocorrelação parcial φ_{kk}^m é diferente de zero para k

menor ou igual a p_m e zero para k maior que p_m . Portanto, a identificação clássica do modelo PAR(p) fundamenta-se em determinar as ordens apropriadas aos operadores autorregressivos de cada período p_m , $m = 1, \dots, s$. Estas ordens são determinadas de acordo com as estimativas $\hat{\phi}_{kk}^m$, $k = 1, \dots, T/4$ e substituindo as autocorrelações pelos respectivos valores amostrais em (11). Se a ordem do operador autorregressivo em um determinado período m for igual a p_m , então $\hat{\phi}_{kk}^m$ terá distribuição aproximadamente normal com média zero e variância $1/T$ quando $k > p_m$. Na ocasião, procura-se a maior ordem i para cada período m de modo que todas as estimativas $\hat{\phi}_{kk}^m$ não sejam mais significativas para $k > i$.

Os parâmetros φ_i^m , $i = 1, \dots, p_m$ são estimados substituindo os parâmetros ρ_j^{m-k} , $j = 0, \dots, (p_m - 1)$, $k = 1, \dots, p_m$ por suas medidas amostrais em (11).

5. EXPERIMENTO COMPUTACIONAL

Nesta seção são apresentados os resultados de um experimento computacional realizado com o objetivo de avaliar o desempenho da aplicação de modelos PAR(p) na modelagem de séries de velocidades de vento e se a performance destes modelos pode ser aprimorada com a suavização prévia das séries temporais por meio da SSA/MSSA. Nos experimentos foram considerados os registros de velocidade do vento provenientes das estações do Instituto Nacional de Meteorologia - INMET (<http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep>) localizadas nas cidades pernambucanas de Petrolina e Pesqueira. Os registros analisados abrangem o período de janeiro de 1996 até dezembro de 2012, onde o período de janeiro de 1996 a dezembro de 2011 foi utilizado como amostra de treinamento (*in sample*) e os 12 últimos meses como amostra de teste (*out of sample*). Inicialmente, procede-se com a filtragem simultânea das séries para em seguida modelar via PAR(p).

A matriz trajetória multidimensional da fase de incorporação MSSA tem dimensão $(2L \times K)$. Neste caso, cada série tem dimensão $T = 192$ e o tamanho de janela utilizado foi $L = 96$, logo a matriz trajetória multidimensional tem dimensão 192×97 . Na Figura 1 são ilustradas as nove primeiras componentes principais e suas respectivas contribuições para as séries de velocidade do vento.

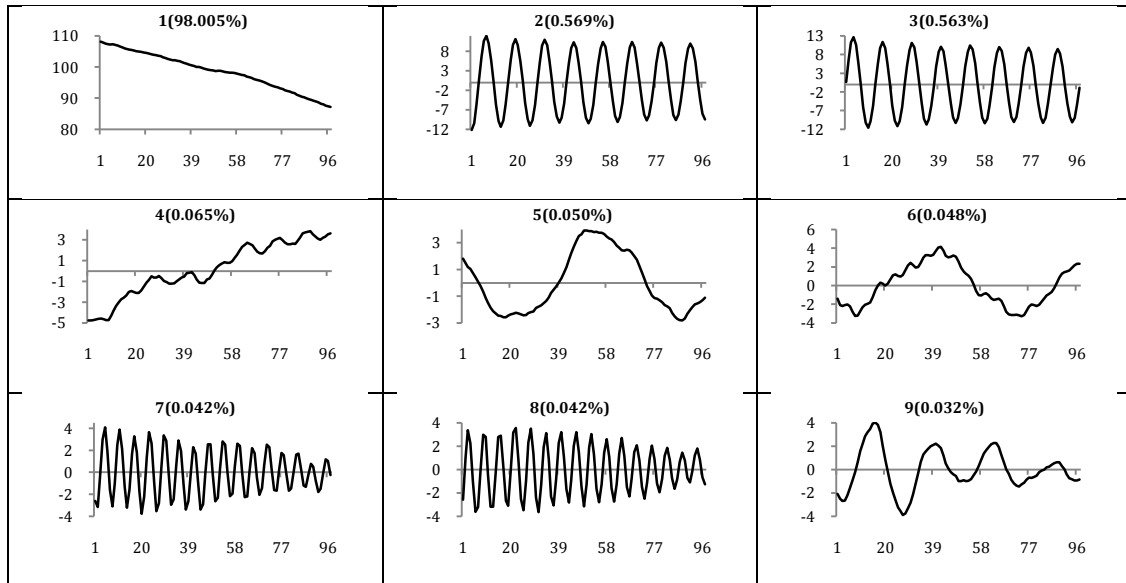


Figura 1. As nove primeiras Componentes Principais da Fase de incorporação MSSA.

Para a filtragem das séries via MSSA, as duas séries são desmembradas nas componentes: Tendência, Harmônica e Ruído. Para a melhor identificação das componentes harmônicas, os gráficos das componentes principais pareadas são utilizados. Na figura 2 são

ilustrados dois pares de componentes principais que fazem parte da componente harmônica. Na Figura 3 são ilustrados dois pares de componentes principais que fazem parte da componente de ruído:

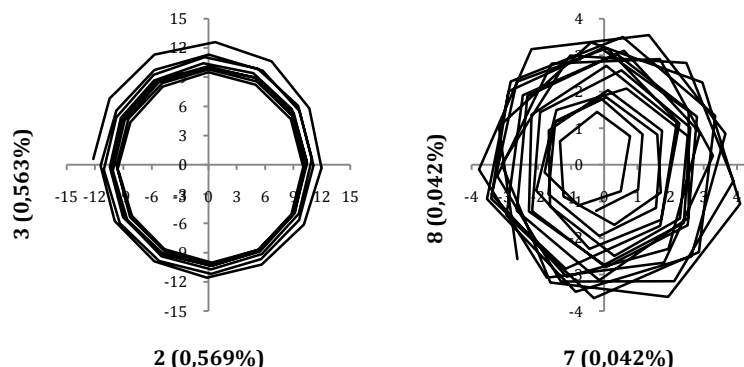


Figura 2. Pares de Componentes Principais pertencentes a componente Harmônica.

Como descrito na seção 2, o número de lados do polígono gerado nos pares da figura 2 determina o período da harmônica. Neste caso, os componentes principais 2 e 3 são harmônicos de período 12, enquanto os componentes principais 7 e 8 são harmônicos de período 6. Os componentes que não são identificados como tendência ou harmônica são considerados ruidosos, conforme ilustrado na Figura 3.

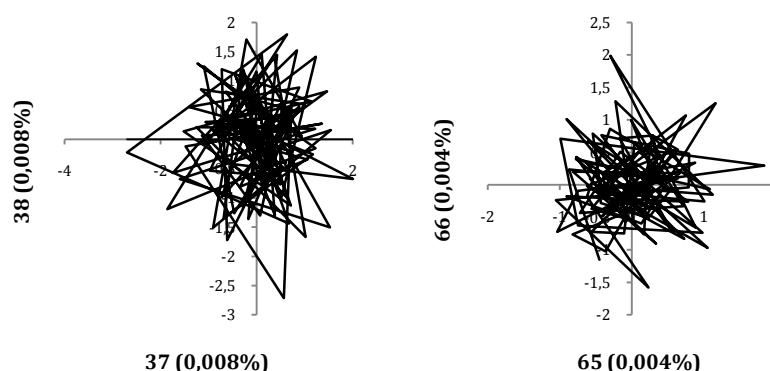


Figura 3. Pares de Componentes Principais pertencentes a componente Ruído.

As correlações ponderadas na Tabela 1 confirmam a separabilidade entre as componentes de tendência, harmônica e ruído.

Tabela 1. Correlação Ponderada entre as componentes.

Componentes	Tendência	Harmônica	Ruído
Tendência	1	0,001	0,003
Harmônica	0,001	1	0,029
Ruído	0,003	0,029	1

Por meio do teste BDS, verificou-se que as componentes ruidosas, mostradas na figura 4, não contem estruturas de autocorrelação (Tabela 2), portanto, elas podem ser removidas a fim de obter as séries aproximadas para a velocidade do vento em cada uma das duas estações anemométricas.

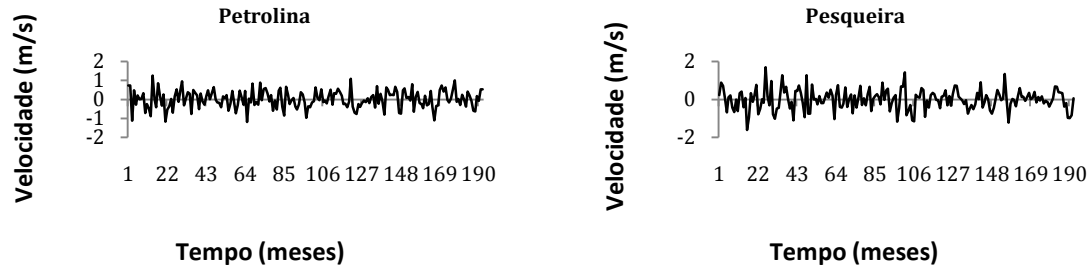


Figura 4. Componente ruidosa das duas estações anemométricas na fase de decomposição MSSA.

A mesma análise feita com a filtragem MSSA foi feita com as séries separadamente via SSA e a modelagem PAR(p) foi feita para os três casos: as séries originais sem a filtragem; as séries com as filtrações SSA individuais e as séries filtradas via MSSA. Para avaliar o poder preditivo dos modelos, as medidas estatísticas de erro MAPE (*Mean Absolute Percentage Error*) dada por:

$$MAPE = \frac{100}{T_h} \sum_{t=1}^{T_h} \frac{|y_t - \hat{y}_t|}{y_t} \quad (12)$$

e RMSE (*Root-Mean-Square Error*) dada por:

$$RMSE = \sqrt{\frac{1}{T_h} \sum_{t=1}^{T_h} (y_t - \hat{y}_t)^2}, \quad (13)$$

onde y_t and \hat{y}_t são respectivamente os valores originais e previstos de velocidade do vento no período t e T_h é o horizonte de previsão, foram calculadas em cada um dos três casos. Os resultados estão apresentados nas tabelas 3 e 4.

Tabela 2. Testes BDS aplicados às séries de ruídos.

	Dim	Estatística BDS	Estatística Z	Erro padrão	p -valor
Petrolina	2	0,006156	1,379423	0,004463	0,1678
	3	0,005718	0,805545	0,007098	0,4205
	4	0,004280	0,506136	0,008457	0,6128
	5	0,004037	0,457835	0,008817	0,6471
	6	0,000435	0,051183	0,008504	0,9592
Pesqueira	2	0,000525	0,925448	0,000568	0,3547
	3	0,001376	1,114620	0,001234	0,2650
	4	0,001597	0,795097	0,002008	0,4266
	5	0,001581	0,553033	0,002859	0,5802
	6	0,005030	1,336379	0,003764	0,1814

Tabela 3. MAPE(%) e RMSE (%) *in sample* dos meses para os modelos PAR(p), PAR(p) - SSA e PAR(p) – MSSA. Estação Petrolina.

PETROLINA	MAPE			RMSE		
Mês	PAR(p)	PAR(p) – SSA	PAR(p) – MSSA	PAR(p)	PAR(p) – SSA	PAR(p) – MSSA
Janeiro	6,1200	3,3072	1,48270	0,3161	0,2150	0,0957
Fevereiro	8,7405	3,1330	0,87904	0,2944	0,2029	0,1070
Março	6,1584	4,1720	0,81247	0,2895	0,2381	0,1178
Abril	5,9539	2,4138	0,63071	0,2665	0,1850	0,1290
Mai	4,4594	3,3287	0,51478	0,2400	0,1867	0,1423
Junho	4,5123	2,1882	0,51592	0,2596	0,1642	0,1567
Julho	2,9755	1,9037	0,54739	0,2274	0,1802	0,1637
Agosto	3,5905	2,3111	0,61316	0,2602	0,2212	0,1556
Setembro	5,0779	2,5236	0,70418	0,2941	0,1977	0,1367
Outubro	5,6350	2,4985	0,73237	0,1778	0,1637	0,1130
Novembro	4,1773	1,9260	0,62408	0,2026	0,1574	0,0941
Dezembro	4,3738	2,5884	0,68437	0,2085	0,1525	0,0877

Tabela 4. MAPE(%) e RMSE (%) *in sample* dos meses para os modelos PAR(p), PAR(p) - SSA e PAR(p) – MSSA. Estação Pesqueira.

PESQUEIRA	MAPE			RMSE		
Mês	PAR(p)	PAR(p) – SSA	PAR(p) – MSSA	PAR(p)	PAR(p) – SSA	PAR(p) – MSSA
Janeiro	6,1182	4,7278	3,5413	0,4879	0,3691	0,3005
Fevereiro	5,1540	2,6814	1,1211	0,7270	0,3308	0,3167
Março	7,1102	1,9755	1,4781	0,3734	0,3443	0,1756
Abril	7,4127	3,1508	1,8216	0,3565	0,3427	0,3281
Mai	5,7543	2,7266	1,7512	0,3714	0,2854	0,2590
Junho	4,3649	2,4026	1,3590	0,3659	0,2773	0,2405
Julho	3,8340	1,9483	1,4165	0,3342	0,2486	0,2301
Agosto	4,5714	2,2650	1,7567	0,3438	0,2145	0,1851
Setembro	5,3439	1,7724	1,4193	0,3830	0,2522	0,1888
Outubro	5,9651	2,1032	1,4086	0,3853	0,2527	0,1777
Novembro	7,0626	2,3336	1,2869	0,4051	0,2572	0,1987
Dezembro	4,4704	0,8727	1,0379	0,2437	0,2336	0,2019

Para a amostra de teste (*out of sample*) foram considerados os 12 últimos meses da amostra e para as duas estações foram calculadas as mesmas estatísticas MAPE e RMSE verificados nas três abordagens. Os resultados são apresentados nas tabelas 5 e 6 para Petrolina e Pesqueira, respectivamente.

Tabela 5. MAPE(%) e RMSE (%) *out of sample* dos meses para os modelos PAR(p), PAR(p) - SSA e PAR(p) – MSSA. Estação Petrolina.

PETROLINA	MAPE			RMSE		
Mês	PAR(p)	PAR(p) – SSA	PAR(p) – MSSA	PAR(p)	PAR(p) – SSA	PAR(p) – MSSA
Janeiro	6,6	4,8	4,4	0,167	0,127	0,107
Fevereiro	10,9	9,5	4,7	0,268	0,231	0,116
Março	13,6	8,8	5,2	0,332	0,218	0,160
Abril	9,5	7,8	5,3	0,311	0,213	0,141
Maio	6,4	5,4	4,8	0,192	0,161	0,140
Junho	11,7	5,8	4,2	0,298	0,174	0,132
Julho	11,3	5,5	3,8	0,301	0,176	0,127
Agosto	12,1	6,0	4,0	0,402	0,206	0,134
Setembro	10,4	6,8	4,2	0,335	0,225	0,137
Outubro	12,4	9,6	4,6	0,343	0,280	0,139
Novembro	10,1	9,3	4,4	0,280	0,254	0,123
Dezembro	8,9	7,3	4,1	0,222	0,188	0,107

Tabela 6. MAPE(%) e RMSE (%) *out of sample* dos meses para os modelos PAR(p), PAR(p) - SSA e PAR(p) – MSSA. Estação Pesqueira.

PESQUEIRA	MAPE			RMSE		
Mês	PAR(p)	PAR(p) – SSA	PAR(p) – MSSA	PAR(p)	PAR(p) – SSA	PAR(p) – MSSA
Janeiro	30,9	18,8	13,7	0,849	0,538	0,387
Fevereiro	25,9	14,1	10,2	0,724	0,377	0,275
Março	18,8	17,1	6,7	0,485	0,454	0,179
Abril	26,4	22,7	19,3	0,662	0,570	0,474
Maio	24,5	23,5	17,4	0,599	0,579	0,432
Junho	27,0	21,6	21,4	0,656	0,535	0,518
Julho	23,7	17,1	14,3	0,591	0,434	0,357
Agosto	25,3	7,2	6,3	0,664	0,196	0,133
Setembro	20,5	8,0	5,6	0,557	0,233	0,167
Outubro	20,3	5,8	4,8	0,584	0,179	0,150
Novembro	28,7	10,9	5,9	0,805	0,334	0,183
Dezembro	7,0	6,4	4,4	0,184	0,133	0,205

6. CONCLUSÕES

Este artigo apresentou a proposta de modelagem PAR(p) com filtragem SSA e MSSA aplicada as séries de velocidade do vento de duas estações anemométricas localizadas no Nordeste do Brasil. Para as filtragens SSA e MSSA, o teste BDS aplicado às séries de ruídos mostram que elas são independentes até a sexta dimensão. Isso garante a princípio que as séries que estão sendo removidas nas filtragens são ruídos e não uma parte do sinal. Outra medida utilizada foi a correlação ponderada entre as componentes de tendência, harmônica e ruído extraídas das séries. As medidas obtidas da correlação ponderada mostram que estas componentes estão bem separadas. Isto faz com que parte do sinal não seja removido como ruído ou parte de ruído não seja utilizado como sinal.

Após as filtragens SSA e MSSA, as séries foram modeladas via PAR(p), tanto as séries originais quando as séries filtradas. Com isso, foram ajustados modelos PAR(p) às séries de velocidade do vento antes e após a filtragem via SSA e MSSA. As medidas MAPE e

RMSE mostram que a filtragem SSA antes da modelagem PAR(p) aumenta a acurácia das previsões e que a filtragem MSSA antes desta modelagem supera os resultados verificados anteriormente com a filtragem SSA. O estudo deste artigo também direciona na possibilidade de usar a modelagem PAR(p) na previsão de velocidade do vento objetivando a inclusão deste tipo de geração de energia no Sistema Energético Brasileiro.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Elshorbagy, A.; Simonovic, S.P.; Panu, U.S. (2002) Noise reduction in chaotic hydrologic time series: facts and doubts, *Journal of Hydrology*, 256, 147-265.
- [2] Golyandina, N., Nekrutkin, V., Zhihgljavsky, A. (2001). *Analysis of time series structure: SSA and related techniques*. Chapman&Hall/CRC. New York, USA.
- [3] Golyandina, N., Stepanov. D. (2005). SSA-based approaches to analysis and forecast of multidimensional time series. *Proceedings of the Fifth Workshop on Simulation*. Department of Mathematics, St. Petersburg State University, Russia. P. 293-298.
- [4] Golyandina, N. (2010). On the choice of parameters in Singular Spectrum Analysis and related subspace-based methods. *Statistics and Its Interface*, 2010, 3, 259-279.
- [5] Hassani, H. (2007) Golyandina, N. (2010). Singular Spectrum Analysis: Methodology and Comparison. *Journal of Data Science*. 5, 239-257.
- [6] Hassani, H. .; Zokaei, M.; Von Rosen, D.; Amiri, S.; Ghodsi, M. (2009). Does noise reduction matter for curve fitting in growth curve models? *Computer Methods and Programs in Biomedicine*, 96, pp. 173-181.
- [7] Hanson, R.T.; Dettinger, M.D.; Newhouse, M.W. (2006). Relations between climatic variability and hydrologic time series from four alluvial basins across the southwestern United States, *Hydrogeology Journal*.
- [8] Hipel, K. W. e McLeod, A. I. (1994) *Time Series Modelling of Water Resources and Environmental Systems*. Amsterdam, The Netherlands: Elsevier.
- [9] Jayawardena, A.W.; Gurung, A.B. (2000) Noise reduction and prediction of hydrometeorological time series: dynamical systems approach vs. stochastic approach, *Journal of Hydrology*, 228, 242-264.
- [10] Maceira, M. E. P.; Penna, D. D. J.; Damázio, J. M. (2006). Geração de cenários sintéticos de energia e Vazão para o planejamento da operação energética, *Cadernos do IME - Série Estatística*, UERJ, Rio de Janeiro, pp. 11-35.
- [10] Schreiber, T.; Grassberger, P. (1991) A simple noise-reduction method for real data, *Phys Letter A*, 160 (5), 411-418.
- [11] Sivakumar, B.; Phoon, K.; Liong, S.; Liaw, C. (1999) A systematic approach to noise reduction in chaotic hydrological time series, *Journal of Hydrology*, 219, pp. 103-135.