

MULTI-CHANNEL SINGULAR SPECTRUM ANALYSIS NA FILTRAGEM DE SÉRIES TEMPORAIS DE ENERGIA NATURAL AFLUENTE

Moisés Lima de Menezes

Universidade Federal Fluminense

Rua Mário Santos Braga, S/N, Campus Valonguinho, Centro, Niterói - RJ, 24.220-900.

moises_lima@msn.com

Reinaldo Castro Souza

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

Universidade do Estado do Rio de Janeiro

Rua São Francisco Xavier, 524, Maracanã, Rio de Janeiro – RJ, 20.550-013.

professorjfm@hotmai.com

Resumo

O Modelo Autorregressivo Periódico da família Box & Jenkins - PAR(p) vem sendo empregado na previsão da Energia Natural Afluyente (ENA), uma variável importante no planejamento da operação energética de sistemas hidrelétricos. Visando aumentar a acurácia das previsões de ENA obtidas pelo modelo PAR(p), propõe-se a aplicação da *Multi-Channel Singular Spectrum Analysis* – MSSA na filtragem simultânea das séries de ENA dos quatro subsistemas que compõem o Sistema Interligado Nacional (SIN). Modelos PAR(p) são ajustados às séries filtradas e as previsões resultantes são comparadas com as obtidas por modelos PAR(p) ajustados às séries originais. Os resultados apresentados no artigo mostram a eficácia da filtragem na melhoria da acurácia das previsões da ENA obtidas pelo PAR(p).

Palavras Chave: filtragem, previsão de séries temporais, energia natural afluyente.

Abstract

The Periodic Autoregressive Model Family Box & Jenkins - PAR(p) has been employed in the prediction of Affluent Natural Energy (ENA), an important variable in the energetic operation planning of hydroelectric systems. In order to increase the accuracy of the ENA's predictions obtained by the PAR(p) model, we propose the application of Multi-Channel Singular Spectrum Analysis - MSSA in the simultaneous filtering of the ENA series of the four subsystems that comprise the National Interconnected System. The PAR(p) models are adjusted to the filtered time series and the resulting predictions are compared with those obtained by models PAR(p) adjusted to the original time series. The results presented in the paper show the effectiveness of filtering in improving the accuracy of ENA's predictions obtained by the PAR(p) model.

Keywords: filtering, time series forecasting, affluent natural energy.

1. INTRODUÇÃO

Em sistemas hidrelétricos a produção energética depende da vazão afluyente às usinas do sistema. Diante da impossibilidade de um conhecimento perfeito das afluências futuras, os agentes responsáveis pela operação do sistema tomam decisões em um ambiente com remarcada incerteza. Vale ressaltar que na cadeia de modelos de apoio a decisão para o planejamento da operação energética considera-se o conteúdo energético das afluências, expresso pela energia natural afluyente (ENA) aos aproveitamentos hidrelétricos. Para mitigar a incerteza, o planejamento da operação energética baseia-se em previsões e séries sintéticas da energia natural afluyente, ambas obtidas por modelos estocásticos.

No caso do setor elétrico brasileiro, a previsão e a geração de séries sintéticas da ENA são realizadas por meio de modelos periódicos autorregressivos de ordem p - PAR(p), uma abordagem adequada em séries temporais sazonais e com estruturas de autocorrelação que dependem das estações do ano, como é o caso das afluências (HIPEL & McLEOD, 1994). As séries temporais da ENA têm discretização mensal e são disponibilizadas para cada um dos quatro subsistemas (Nordeste, Norte, Sudeste/Centro-Oeste e Sul) que compõem o Sistema Interligado Nacional (SIN).

Tradicionalmente, uma série temporal pode ser expressa como a soma das componentes de tendência, sazonalidade/ciclo e ruído. A remoção da componente ruído pode contribuir para a melhor identificação do processo estocástico subjacente a série temporal e, conseqüentemente, melhorar a acurácia das previsões. SCHREIBER & GRASSBERGER (1991) e posteriormente SIVAKUMAR (1999) e JAYAWARDENA & GURUNG (2000) mostram que a remoção de ruídos em séries hidrológicas melhora a acurácia das previsões. Uma boa revisão do assunto é apresentada por ELSHORBAGY et al (2002).

Uma técnica que tem sido utilizada com sucesso na remoção de ruídos é a análise espectral de valores singulares (GOLYANDINA et al., 2001; HASSANI, 2007; HASSANI et al 2009) ou SSA (*Singular Spectrum Analysis*). Por meio da SSA pode-se decompor uma série temporal em duas componentes: o sinal propriamente dito e o ruído. A componente de ruído é descartada, enquanto o sinal é a série temporal considerada na especificação e ajuste de um modelo de previsão.

Com base nas considerações anteriores, o presente artigo apresenta os resultados de um experimento computacional conduzido com o objetivo de avaliar o efeito da remoção dos ruídos das séries de ENA sobre a acurácia das previsões geradas por modelos PAR(p). As séries de ENA dos quatro subsistemas foram filtradas por meio da *Multi-channel Singular Spectrum Analysis* - MSSA (GOLYANDINA et al., 2005), uma extensão multivariada da SSA que permite a filtragem conjunta de duas ou mais séries temporais..

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 ENA antes e após a filtragem por MSSA. 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 *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.

Tome a sequência $\{E_l\}_{l=1}^d$ de matrizes elementares na SVD, em (2). Agrupe-as em $m \leq d$ grupos *disjuntos* e assuma 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 $\{E_l\}_{l=1}^d = \cup_{i=1}^m \{X_{I_{ij}}\}_{j=1}^{p_i}$, onde $m \leq d$. 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 $\cup_{i=1}^m \{X_{I_{ij}}\}_{j=1}^{p_i}$ 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). O objetivo do agrupamento é diminuir o número de componentes (ou matrizes elementares) na SVD da matriz trajetória X . 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 $\cup_{i=1}^m \{X_{I_{ij}}\}_{j=1}^{p_i}$ de matrizes elementares na SVD. Dado que F é um mapa invertível, segue que F^{-1} sobre a matriz trajetória X é tal que:

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

A componente $[y_t^{(i)}]_{1 \times T}$ 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} .

Tome a matriz trajetória X e assuma 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 $[y_t^{(i)}]_{1 \times T}$ 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 $[y_t^{(i)}]_{1 \times T}$ 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*, *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

$$\rho_{ij}^{(w)} = \frac{(Y_t^{(i)}, Y_t^{(j)})_w}{\|Y_t^{(i)}\|_w \|Y_t^{(j)}\|_w}, \text{ onde } \| \cdot \| \text{ é a norma euclidiana, } (\cdot)_w \text{ é o produto interno tal que:}$$

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

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 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

A 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 sistema de séries temporais em face de uma única série.

Considere o sistema de s séries temporais de tamanho T

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

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 $X_j^{(k)} = (y_j^{(k)}, \dots, y_{j+L-1}^{(k)})^T$, $j = 1, \dots, K$ 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).

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

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:

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

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 X pode ser reescrita no que pode ser chamado de 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 $p = (p_1, p_2, \dots, p_s)$, onde 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 desta metodologia é o 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 $\hat{\varphi}_{kk}^m$, $k = 1, \dots, T/4$ 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 $\hat{\varphi}_{kk}^m$ 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 $\hat{\varphi}_{kk}^m$ 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

Inicialmente, procede-se a filtragem simultânea das séries mensais de ENA dos quatro subsistemas, disponíveis no Operador Nacional do Sistema Elétrico (ONS) e referentes ao período de janeiro de 1931 a dezembro de 2012.

A matriz trajetória multidimensional da fase de incorporação MSSA tem dimensão $(4L \times K)$. Neste caso, cada série tem dimensão $T = 984$ e o tamanho de janela utilizado foi $L = 492$, logo a matriz trajetória multidimensional tem dimensão 1968×493 . Na Figura 1 são ilustradas as oito primeiras componentes principais e suas respectivas contribuições para as séries de ENA:

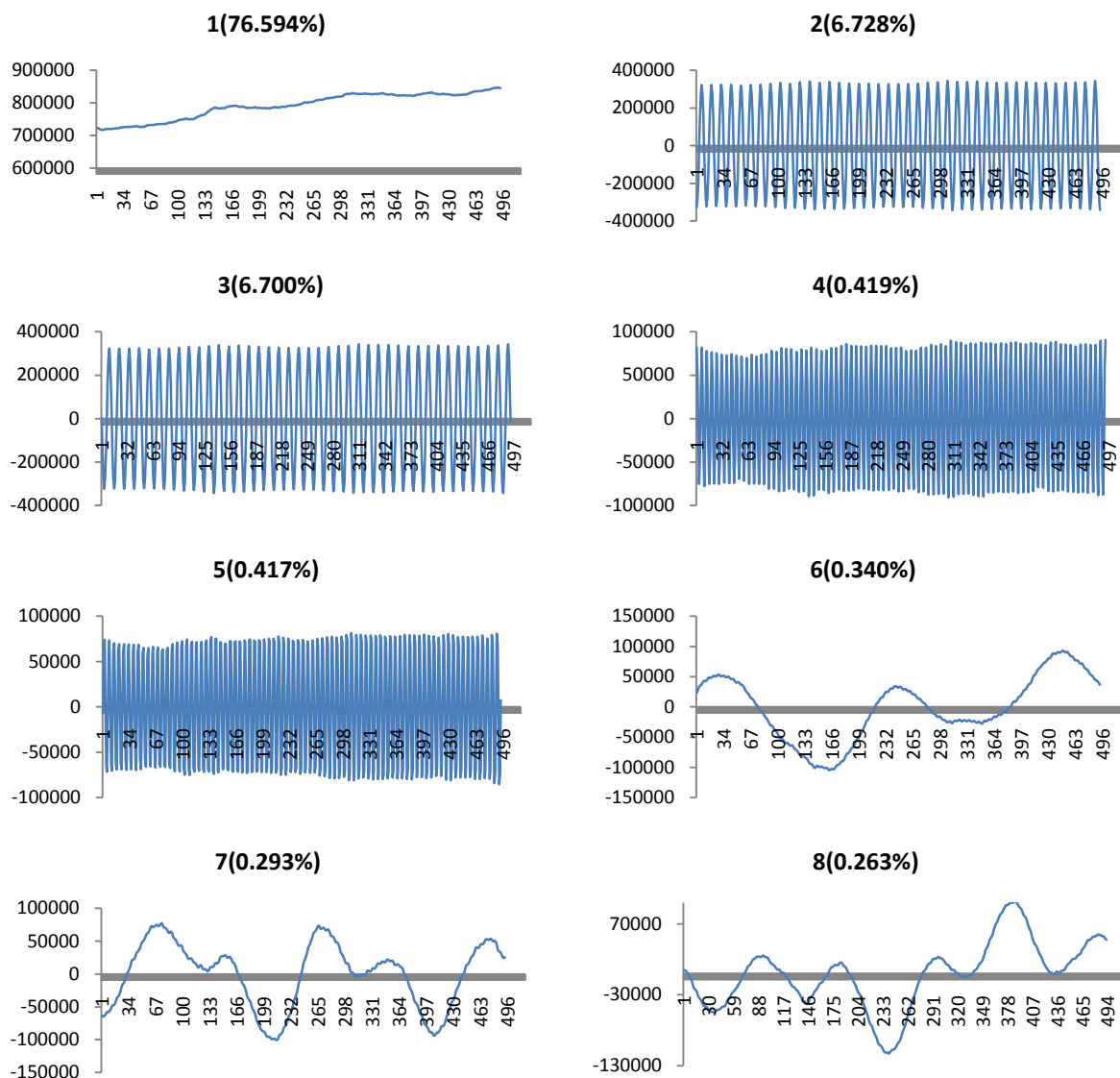


Figura 1. As oito primeiras Componentes Principais da Fase de incorporação MSSA.

Para a filtragem das séries via MSSA, as quatro 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 três pares de componentes principais que fazem parte da componente de ruído:

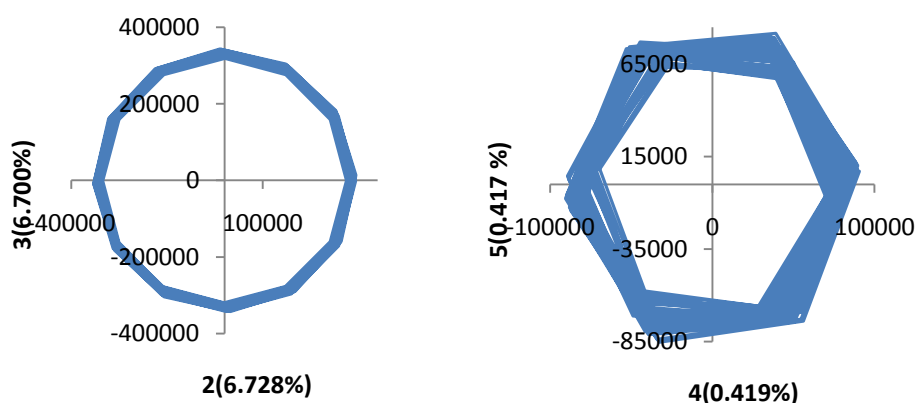


Figura 2. Pares de Componentes Principais pertencentes a componente Harmônica.

Como descrito na seção 1.4, 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 4 e 5 são harmônicos de período 6. Os componentes que não são identificados como tendência ou harmônica são considerados ruidosos, conforme ilustrado na Figura 3.

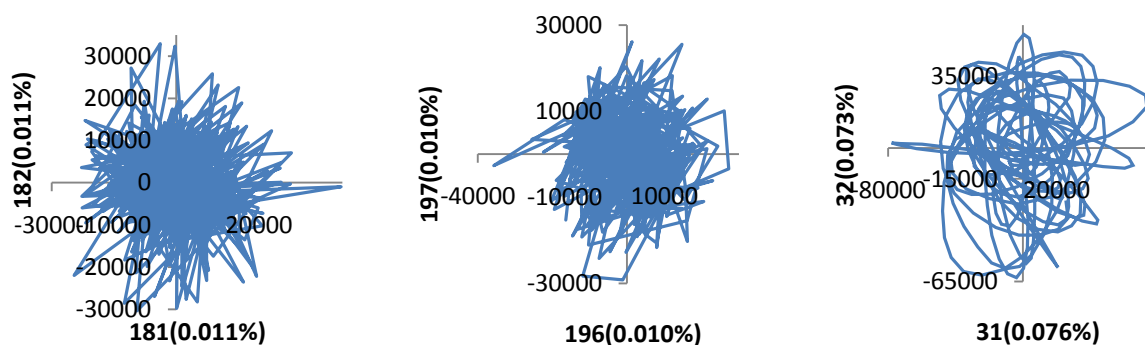


Figura 3. Pares de Componentes Principais pertencentes a componente Ruído.

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 | 0,011 |
| Harmônica | 0,001 | 1 | 0,042 |
| Ruído | 0,011 | 0,042 | 1 |

Por meio do teste BDS, verificou-se que as componentes ruidosas não contêm estruturas de autocorrelação (Tabela 2), portanto, elas podem ser removidas a fim de obter as séries aproximadas para a ENA em cada um dos quatro subsistemas na Figura 4.

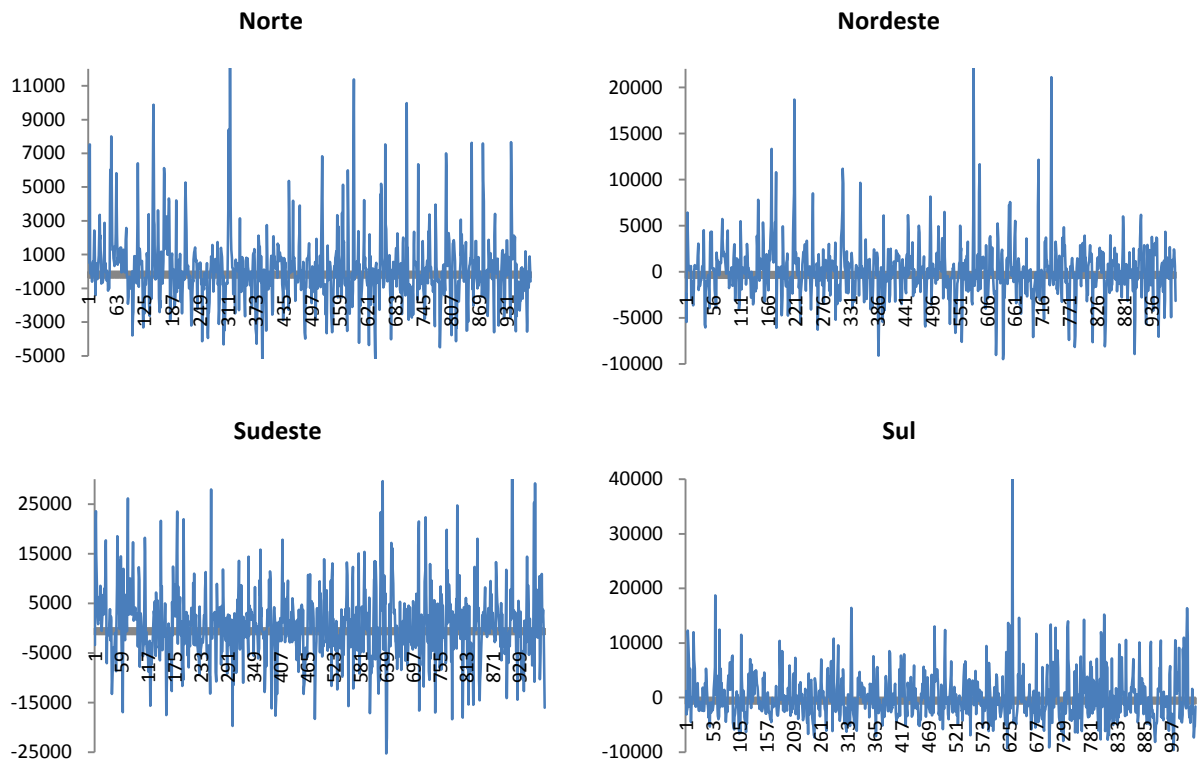


Figura 4. Componente ruidosa dos quatro subsistemas na fase de decomposição MSSA.

Tabela 2. Testes BDS aplicados às séries de ruídos.

| | Dim. | Estatística BDS | Estatística Z | Probab. | | Dim. | Estatística BDS | Estatística Z | Probab. |
|---------|------|-----------------|---------------|---------|----------|------|-----------------|---------------|---------|
| Norte | 2 | 1,79E-07 | 0,012852 | 0,9697 | Nordeste | 2 | 3,09E-06 | -0,132011 | 0,8950 |
| | 3 | 5,30E-07 | 0,017004 | 0,9864 | | 3 | 5,24E-06 | 0,100331 | 0,9201 |
| | 4 | 1,04E-06 | 0,020036 | 0,9840 | | 4 | 1,46E-05 | 0,167668 | 0,8668 |
| | 5 | 1,71E-06 | 0,022483 | 0,9821 | | 5 | 2,51E-05 | 0,196610 | 0,8441 |
| | 6 | -9,92E-06 | -0,096179 | 0,9234 | | 6 | 3,45E-05 | 0,199968 | 0,8415 |
| Sudeste | 2 | -4,13E-09 | -1,060615 | 0,2889 | Sul | 2 | -1,41E-06 | -0,025930 | 0,9793 |
| | 3 | -2,67E-13 | -0,028212 | 0,9775 | | 3 | -6,29E-06 | -0,051989 | 0,9585 |
| | 4 | -1,73E-17 | -0,001006 | 0,9992 | | 4 | -1,26E-05 | -0,062260 | 0,9504 |
| | 5 | -1,13E-21 | -4,12E-05 | 0,9999 | | 5 | -2,03E-05 | -0,068674 | 0,9452 |
| | 6 | -7,37E-26 | -1,83E-06 | 0,9999 | | 6 | -3,15E-05 | -0,078835 | 0,9372 |

Os modelos PAR(p) foram ajustados às séries filtradas e às séries originais e na sequência foram obtidas previsões segundo os dois ajustes. Na Tabela 3 são apresentados os valores do MAPE para cada mês do horizonte de previsão, em cada uma das modelagens.

Tabela 3. MAPE ao longo do horizonte de previsão segundo cada modelagem

| MAPE | Norte | | Nordeste | | Sudeste | | Sul | |
|------|--------|---------------------|----------|---------------------|---------|---------------------|--------|---------------------|
| | PAR(p) | PAR(p) – MSSA | PAR(p) | PAR(p) – MSSA | PAR(p) | PAR(p) – MSSA | PAR(p) | PAR(p) – MSSA |
| Jan | 21.315 | 4.0486 | 20.087 | 4.0741 | 19.099 | 3.5581 | 43.863 | 6.1534 |
| Fev | 22.487 | 3.2483 | 29.174 | 4.4951 | 20.425 | 3.7088 | 42.421 | 5.4864 |
| Mar | 14.009 | 2.0893 | 24.617 | 4.4224 | 17.947 | 3.2184 | 31.303 | 8.8273 |
| Abr | 14.146 | 2.3321 | 29.199 | 5.0078 | 12.334 | 3.2942 | 45.127 | 14.769 |
| Mai | 12.960 | 2.8843 | 14.683 | 6.9024 | 10.379 | 4.6633 | 79.801 | 8.0755 |
| Jun | 10.511 | 5.5560 | 7.2714 | 10.927 | 10.441 | 5.1704 | 55.224 | 5.1090 |
| Jul | 6.8268 | 4.9606 | 5.1856 | 12.670 | 8.0073 | 5.5021 | 46.304 | 5.1961 |
| Ago | 5.7416 | 10.351 | 4.5223 | 12.346 | 9.2411 | 7.1534 | 67.320 | 4.3921 |
| Set | 8.3045 | 21.175 | 6.7781 | 11.936 | 14.192 | 8.0904 | 51.134 | 4.2216 |
| Out | 13.292 | 9.5091 | 13.626 | 5.9950 | 18.253 | 5.9805 | 48.869 | 3.2251 |
| Nov | 19.925 | 5.7945 | 24.585 | 7.5249 | 13.677 | 4.5263 | 42.127 | 4.4839 |
| Dez | 22.547 | 5.5226 | 25.193 | 4.2901 | 15.421 | 4.6130 | 46.571 | 7.2572 |

6. CONCLUSÕES

Este artigo apresentou a proposta de modelagem PAR(p) – MSSA aplicada às séries de ENA dos quatro subsistemas Norte, Nordeste, Sudeste/Centro-Oeste e Sul. Para a filtragem 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 na filtragem MSSA 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 a filtragem MSSA, as séries foram modeladas via PAR(p), tanto as séries originais quando as séries filtradas MSSA. Com isso, foram ajustados modelos PAR(p) às séries de ENA antes e após a filtragem via MSSA. As medidas MAPE mostram que a modelagem PAR(p) – MSSA aumenta a acurácia das previsões, por serem menores, em sua maioria, que as medidas MAPE obtidas na modelagem PAR(p) com os dados originais.

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 reletade 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 systemas approach vs. stochastic approach, *Journal of Hydrology*, 228, 242-264.
- [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.; Liang, S.; Liaw, C. (1999) A systematic approach to noise reduction in chaotic hydrological time series, *Journal of Hydrology*, 219, pp. 103-135.