



## ALGUMAS ABORDAGENS PARA IDENTIFICAÇÃO DE RUÍDOS NA ANÁLISE ESPECTRAL SINGULAR DE SÉRIES TEMPORAIS

**Keila Mara Cassiano**

Universidade Federal Fluminense – UFF  
Rua Mário Santos Braga S/N, Campus Valonguinho, Centro - Niterói, Brasil.  
[keilamath@hotmail.com](mailto:keilamath@hotmail.com)

**Moisés Lima de Menezes**

Universidade Federal Fluminense – UFF  
Rua Mário Santos Braga S/N, Campus Valonguinho, Centro - Niterói, Brasil.  
[moises\\_lima@msn.com](mailto:moises_lima@msn.com)

**José Francisco Moreira Pessanha**

Universidade do Estado do Rio de Janeiro – UERJ  
Av. Maracanã, Tijuca - Rio de Janeiro, Brasil.  
[professorjfm@hotmail.com](mailto:professorjfm@hotmail.com)

### RESUMO

O objetivo deste artigo é apresentar diferentes métodos para remoção de ruídos de séries temporais com o uso da Análise Espectral Singular (SSA). Para este propósito foram utilizadas três abordagens, na fase de agrupamento do método SSA: análise de componentes principais (ACP), análise de agrupamentos integrada com ACP e análise gráfica dos vetores singulares. Adicionalmente, testes estatísticos foram realizados a fim de se obter evidências empíricas da existência de independência estatística e estacionariedade de segunda ordem na série temporal de ruídos removidos. Para ilustrar a aplicação dos métodos, considerou-se a série temporal de Vazão da Usina Hidrelétrica Governador Bento Munhoz, localizada na Bacia do Rio Paraná, Brasil.

**PALAVRAS CHAVE. Remoção de Ruídos, SSA, Séries Temporais.**

**Área principal (EST)**

### ABSTRACT

The aim of this paper is to present different methods to remove noise from time series using the Singular Spectrum Analysis (SSA). For this purpose, three approaches were used in the grouping step of the method SSA: Principal Component Analysis (PCA), Clustering Analysis integrated with PCA and Graphical Analysis of Singular Eigenvectors. In addition, statistical tests were performed in order to empirically demonstrate statistical independence and second-order stationarity in the time series of noise removed. To illustrate the application of methods, we considered the time series of flow of the Governor Bento Munhoz Hydroelectric Plant, located on the Paraná River Basin.

**Keywords: Remotion of Noise, SSA, Time Series.**

**Main area (EST)**

## 1. Introdução

Análise Espectral Singular (SSA - *Singular Spectrum Analysis*) é um recente e poderoso método em séries temporais e 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 a economia e matemática financeira, da meteorologia e oceanografia a ciências sociais (Hassani, 2007).

O objetivo 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 deste artigo, 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, uma matriz passível de ser 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. 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, 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.

Nesta perspectiva, são apresentadas três diferentes abordagens para remoção de ruídos de séries de tempo com o uso do método SSA: análise de componentes principais (ACP), análise de agrupamentos integrada com ACP e análise gráfica dos vetores singulares. Adicionalmente, foram realizados testes estatísticos sob a série temporal de resíduos extraída de cada abordagem, a fim de garantir, estatisticamente, a existência de independência e estacionariedade de segunda ordem (Hamilton, 1994) - que são atributos das séries temporais de ruídos.

Para a ilustração das três abordagens, considerou-se a série temporal de Vazão da Usina Hidrelétrica Governador Bento Munhoz, localizada na Bacia do Rio Paraná, em Pinhão, estado do Paraná, Brasil. A escolha de uma série de vazão deve-se a importância da previsão para o Sistema Elétrico Brasileiro (SEB). Uma das principais características do SEB reside no fato da sua capacidade de geração ser predominantemente hidráulica. Tal característica contribui substancialmente para que a matriz energética brasileira seja considerada uma das mais renováveis do mundo. Não obstante, devido às incertezas nos regimes das vazões naturais (Terry et al., 1986), o SEB está submetido a um significativo risco hidrológico. Para mitigá-lo, o SEB conta com usinas termelétricas que complementam a geração hidroelétrica no país. Adicionalmente, o sistema ainda dispõe de hidroelétricas com reservatórios de grande capacidade de acumulação que permitem a regularização plurianual e, desta forma, protegem a geração de energia elétrica dos efeitos decorrentes de longos períodos secos. Desse modo, para que estes recursos sejam utilizados de forma sustentável e contribuam à economicidade e à segurança do fornecimento de energia elétrica, a operação do parque termelétrico e das usinas hidroelétricas deve ser realizada de forma coordenada, a fim de se alcançar um equilíbrio entre o melhor uso da água e a minimização das despesas com combustíveis nas unidades termelétricas. Tal coordenação é realizada por meio de uma cadeia de modelos de otimização e simulação que suportam os processos de tomada de decisão no planejamento e na programação da operação do Sistema Interligado Nacional – SIN (Terry et al., 1986). Os resultados destes modelos são sensíveis às previsões das séries temporais de vazões nos aproveitamentos hidroelétricos. No entanto, estas séries temporais são caracterizadas por acentuada sazonalidade e incertezas (erros e imprecisões) na sua mensuração e uma

forma factível de lidar com tais sazonalidades e imprecisões (presentes nas séries de vazões) é o uso de SSA para remoção dos ruídos das séries.

O artigo está organizado em seis seções. Na seção 2, tem-se uma breve apresentação do método SSA. Na seção 3, são descritos os métodos para seleção das componentes, responsáveis pela separação das componentes SSA entre as categorias sinal e ruído. Os dados relativos ao estudo de caso são descritos na seção 4. Os principais resultados obtidos são apresentados na seção 5. Por fim, as conclusões são apresentadas na seção 6.

## 2. Análise Espectral Singular

A *Análise Espectral Singular* (SSA) é um método de processamento de sinais que pode ser utilizado, dentre outras aplicações, na *remoção de ruído de séries de tempo* (Golyandina et al., 2001). A versão básica do método SSA pode ser dividida em duas etapas: *decomposição* e *reconstrução*.

### 2.1. Decomposição

A etapa da *decomposição* pode ser subdividida em duas subetapas: *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).

Considere o *operador normal e compacto*  $S := XX^T$  (Kubrusly, 2001). 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 parcial:  $0 \geq \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L$ .

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 *decomposição em valores singulares*, em (1).

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

Onde:  $X_l := \lambda_l^{1/2} U_l V_l^T$  e os conjuntos  $\{\lambda_l^{1/2}\}_{l=1}^L$  e  $\{U_l\}_{l=1}^L$  são, respectivamente, de nominados por *espectro singular* e de *vetores singulares* da matriz trajetória  $X$ . A coleção  $(\lambda_l, U_l, V_l)$  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:

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

### 2.2. Reconstrução

A etapa de *reconstrução* pode ser subdividida em duas subetapas: *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  $\{X_l\}_{l=1}^d$  de matrizes elementares na SVD, em (2). Agrupe-as em  $m \leq d$  grupos *disjuntos* utilizando algum método [por exemplo, com o auxílio da *análise de componentes principais* [Seção 3.1], ou *análise de agrupamentos integrada com ACP* [Seção 3.2], ou *agrupamento gráfica de vetores singulares* [Seção 3.3] 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  $\{X_l\}_{l=1}^d = \cup_{i=1}^m \{X_{l_{ij}}\}_{j=1}^{p_i}$ , onde  $m \leq d$ . A matriz elementar  $X_{l_i}$  gerada a partir do grupo  $\{X_{l_{ij}}\}_{j=1}^{p_i}$  é dada por  $X_{l_i} = \sum_{j=1}^{p_i} X_{l_{ij}}$ , de modo que a identidade em (2) pode ser reescrita como em (3).

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

É importante salientar que o procedimento de agrupamento pode ser realizado também sob a sequência  $\cup_{i=1}^m \{X_{l_{ij}}\}_{j=1}^{p_i}$  de igual forma ao realizado na sequência  $\{X_l\}_{l=1}^d$  utilizando, inclusive, um método ou critério de agrupamento diferente do aplicado sob a sequência  $\{X_l\}_{l=1}^d$  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  $X_{l_i}$  pode ser mensurada pela razão de valores singulares dada por  $\sum_{j=1}^{p_i} (\lambda_{l_{ij}})^{1/2} / \sum_{l=1}^d (\lambda_l)^{1/2}$ . Tome o conjunto  $\cup_{i=1}^m \{X_{l_{ij}}\}_{j=1}^{p_i}$  de matrizes elementares na SVD. Dado que  $F$  é um mapa invertível, segue que  $F^{-1}$  sobre a matriz trajetória  $X$  é tal que:

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

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  $X_{l_i}$ .

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  $X_{l_i}$ . O elemento  $y_t^{(i)}$  da componente SSA  $\left[y_t^{(i)}\right]_{1 \times T}$  é calculado por meio da *média diagonal*, que é definida em (4), a partir da matriz elementar  $X_{l_i}$ .

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

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  $\sum_{j=1}^{p_i} (\lambda_{l_{ij}})^{1/2} / \sum_{l=1}^d (\lambda_l)^{1/2}$

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* (ciclo e sazonalidade) e *ruído*.

### 2.3. Separabilidade

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:  $\|Y_t^{(i)}\|_w = \sqrt{(Y_t^{(i)}, Y_t^{(i)})_w}$  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).

### 3. Métodos de Remoção de Ruído

No contexto de séries de tempo, os *métodos de remoção de ruído* consistem, basicamente, em se gerar uma série temporal aproximada  $[\tilde{y}_t]_{1 \times T}$  que seja menos ruidosa que a série temporal original  $[y_t]_{1 \times T}$ . Neste contexto, tem-se que qualquer série temporal pode ser decomposta em duas componentes: uma com estrutura de dependência temporal (*linear* ou *não linear*) e outra sem qualquer estrutura de dependência no tempo (que é conhecida como série temporal de ruídos). Tal decomposição é dada em (5).

$$[y_t]_{1 \times T} = [\tilde{y}_t]_{1 \times T} + [\varepsilon_t]_{1 \times T} \quad (5)$$

Onde:  $[y_t]_{1 \times T}$  é a série temporal original;  $[\tilde{y}_t]_{1 \times T}$  é a série temporal aproximada; e  $[\varepsilon_t]_{1 \times T}$  é o ruído. Neste artigo, são utilizados três métodos com esta finalidade: análise de componentes principais (ACP), análise de agrupamentos integrada com ACP e análise gráfica dos vetores singulares.

#### 3.1. Análise de Componentes Principais

Cada valor singular  $\lambda_l$  resultante da SVD quantifica a energia da matriz trajetória  $X$  que está contida na matriz elementar  $X_l$ . Seja  $X = \sum_{i=1}^m X_{l_i}$ . Por meio da *análise de componentes principais* (ACP) (Vasconcelos, 2011), pretende-se determinar um valor ótimo  $l'$  no conjunto de índices  $\{1, \dots, m\}$ , de tal forma que a série temporal  $[y_t]_{1 \times T} - \sum_{i=1}^{l'} [y_t^{(i)}]_{1 \times T}$  seja estatisticamente classificada como ruído.

#### 3.2. Análise de Agrupamentos

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 Aldenfender e 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. Neste artigo, foi o utilizado o *método de agrupamento hierárquico* (Aldenfender, Blashfield, 1984), em virtude da sua simplicidade. 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.

No estudo de caso, foi adotado o método de *agrupamento hierárquico de encadeamento simples* ou (*single-linkage*) para se dividir o conjunto de vetores obtidos pela expansão SVD em três grupos excludentes. O algoritmo utilizado foi o *aglomerativo* (Aldenfender, Blashfield, 1984), ou seja, o algoritmo inicia com  $N$  *clusters*, cada um contendo apenas um vetor, e estes são sucessivamente fundidos dois a dois por meio de um procedimento iterativo até que restem apenas dois *clusters*. Em cada estágio do processo de aglomeração, o conjunto de  $N$  objetos é agrupado em um determinado número de grupos e a distância entre estes é calculada (Aldenfender e Blashfield, 1984).

#### 3.3. 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  $\omega$ , fase igual a  $\delta$ , amplitude igual a  $\xi$  e período  $\rho = \frac{1}{\omega}$  definido como um divisor do tamanho da janela  $L$  e  $K$ . Se o parâmetro  $\rho$  assume um

valor inteiro, então  $\rho$  é classificado como *período do harmônico* (Morettin, Toloí, 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 exhibe um padrão circular (Morettin, Toloí, 2006). Por sua vez, se  $\rho = \frac{1}{\omega}$  é um inteiro, então o diagrama de dispersão exhibe um polígono regular com  $\rho$  vértices. Para uma frequência  $\omega = \frac{m}{n} < 0.5$  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.

#### 4. Estudo de Caso

Localizada na Bacia do Rio Paraná, a Usina Hidrelétrica Governador Bento Munhoz da Rocha Neto (UHE GB Munhoz) foi construída ao longo do Rio Iguazu, no município de Pinhão, e fica a 5 km da jusante da foz do Rio Areia e a 240 km de Curitiba. Foi construída no período entre 1975 e 1980 e é a maior usina da Companhia Paranaense de Energia, com capacidade de geração de 1.676 MW, distribuída em quatro unidades geradoras de 419 MW de potência.

A barragem da usina é de enrocamento compactado com face de concreto, tendo 160 m de altura e 828 m de comprimento. O vertedouro do tipo calha e escavado a céu aberto foi projetado para uma descarga máxima de 11.000 m<sup>3</sup>/s (COPEL, 2012).

Dada a relevância da UHE GB Munhoz, foi utilizada, neste artigo, a série temporal mensal das *médias diárias* de vazão da UHE GB Munhoz, com cardinalidade igual a 924 meses (de janeiro de 1931 a dezembro de 2007) e apresentada na figura 1.

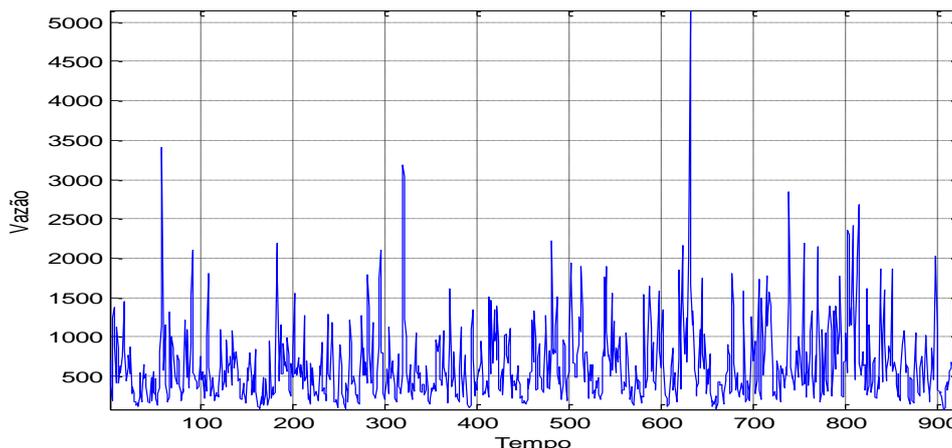


Figura 1 - Série Temporal de Vazão (m<sup>3</sup>/s) da UHE GB Munhoz.

#### 5. Resultados e Discussões

Os métodos de análise de componentes principais (ACP), de análise de agrupamento integrada com ACP e de análise gráfica de vetores singulares foram utilizados na fase de agrupamento do método SSA, com a finalidade de se remover as matrizes elementares que geram componentes SSA estatisticamente classificadas com ruídos pelos testes BDS (Brock, Dechert, Scheinkman, 1987) e *Ljung-Box* (Morettin, Toloí, 2006).

##### 5.1. Aplicação da Análise de Componentes Principais

A análise de componentes principais (ACP) sobre SVD foi implementada no *software* MATLAB. O valor ótimo para o parâmetro  $L$  (tamanho da janela na fase de decomposição do método SSA) foi igual a 362 (ou seja, o número de vetores singulares na SVD foi igual a 362). O valor ótimo de truncamento  $N$  foi igual a 201 (ou seja, dos 362 vetores singulares na SVD, foram mantidos na SVD somente os 201 primeiros vetores singulares; enquanto que os outros foram classificados como ruído e removidos). O método utilizado para obtenção dos valores ótimos dos parâmetros  $L$  e  $N$  foi o método de tentativa e erro. Na Figura 2, tem-se os gráficos sobrepostos da série temporal de vazão da UHE GB

Munhoz (original) e sua aproximação (ou versão filtrada) através do método de ACP na SVD. Note que parte da energia (classificada estatisticamente como ruído) foi removida por ACP.

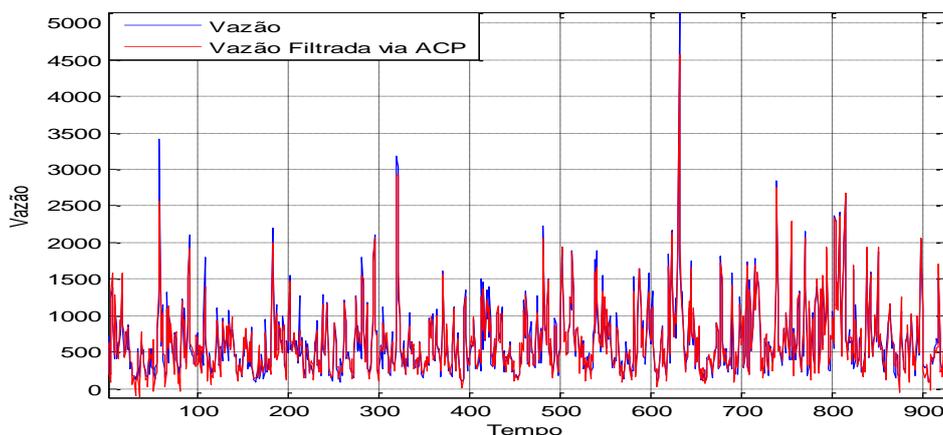


Figura 2 - Série Temporal de Vazão da UHE GB Munhoz e sua Versão Aproximada através da ACP.

### 5.2. Aplicação da Análise de Agrupamentos Integrados com ACP

A ACP foi implementada no *software* MATLAB. Os valores ótimos para os parâmetros  $L$  e  $N$  foram, respectivamente, iguais a 362 e 201 (os mesmos valores da Seção 5.1). Após a ACP, foram tomados os vetores singulares remanescentes na SVD e agrupados, por meio do método do agrupamento *hierárquico* (referenciado na Seção 3.3), em 3 *clusters* (*grupos*). Em cada *cluster* foi gerada uma componente, portanto, a série temporal da vazão da UHE GB Munhoz foi decomposta em 3 componentes. Os testes estatísticos BDS e *Ljung-Box* foram aplicados às componentes SSA e verificou-se que a componente SSA 3 (oriunda do *cluster* 3) possui propriedades estatísticas de ruído. A análise de agrupamento foi implementada no *software* R, com o uso do pacote *Rssa*, e os testes estatísticos, no *software* *EViews*. Na Figura 3, têm-se as componentes SSA obtidas a partir dos 3 *clusters*.

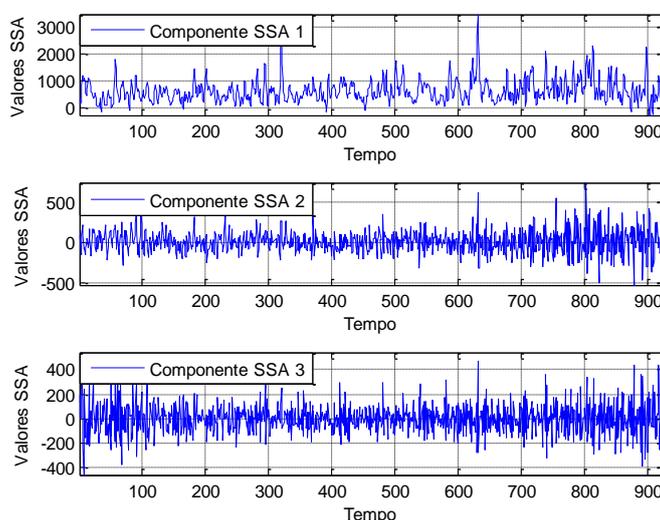
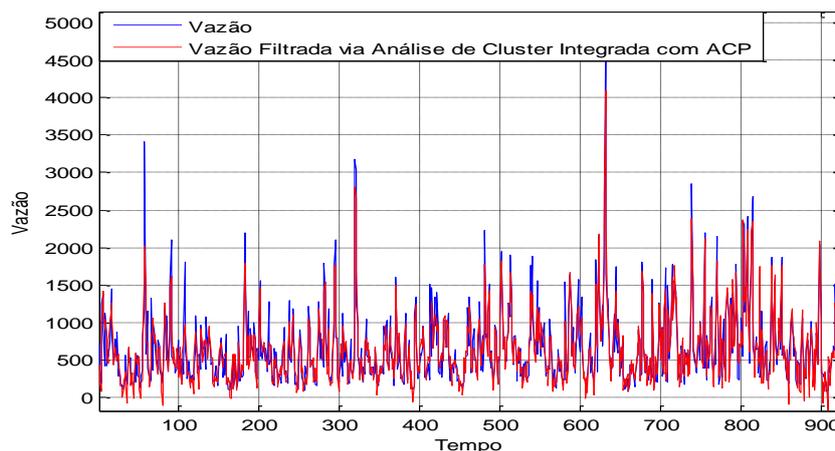


Figura 3 - Componentes SSA para a Série Temporal de Vazão da UHE GB Munhoz Aproximada via ACP.

Foi descartada a componente SSA associada ao *cluster* 3, visto que possui propriedades estatísticas de ruído, gerando uma aproximação da série temporal de vazão da UHE GB Munhoz.

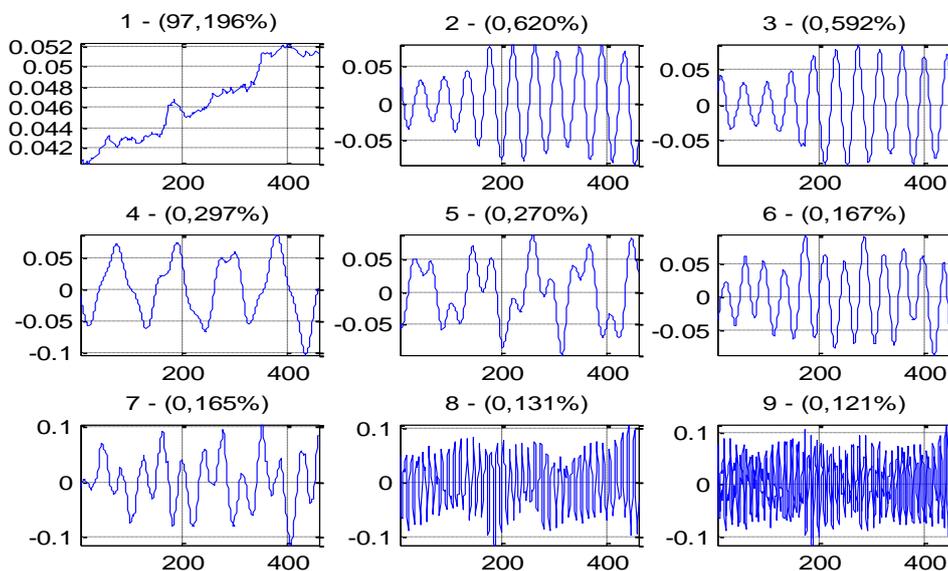


**Figura 4** - Série Temporal de Vazão da UHE GB Munhoz e sua Versão Filtrada através da Análise de Agrupamento Integrada com ACP.

Na Figura 4, pode-se visualizar que parte da energia (classificada estatisticamente como ruído) foi removida por meio da análise de agrupamento integrada com ACP.

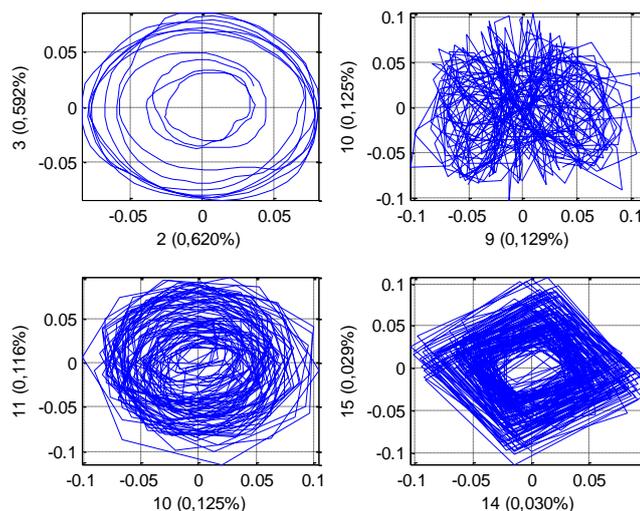
### 5.3. Aplicação da Análise Gráfica dos Vetores Singulares

Na abordagem de análise gráfica de vetores singulares na SVD, foi utilizado o valor ótimo para o parâmetro  $L$  foi igual a 462. Na Figura 5, têm-se os 9 primeiros (e principais) vetores singulares na SVD da matriz trajetória da série temporal de Vazão da UHE GB Munhoz. Na primeira linha, da direita para a esquerda, têm-se os vetores singulares 1, 2 e 3. E assim sucessivamente. Esta abordagem foi implementada no *software Caterpillar SSA* (GISTATGROUP, 2010) e no MATLAB.



**Figura 5** - 9 Primeiros Vetores Singulares na SVD da Matriz Trajetória da Série Temporal de Vazão da UHE GB Munhoz.

Na figura 6, têm-se os gráficos de dispersão de alguns pares de vetores singulares. Por meio de sua análise, pode-se verificar se determinado par possui associação de forma linear ou não linear, bem como se são classificadas como componente de tendência, harmônica ou de ruído (Golyandina et al., 2001, Hassani, 2007, Morettin, Toloi, 2006).



**Figura 6** - Gráficos de Dispersão dos Pares de Autovalores da SVD: (2,3); (9,10); (10,11); e (14,15).

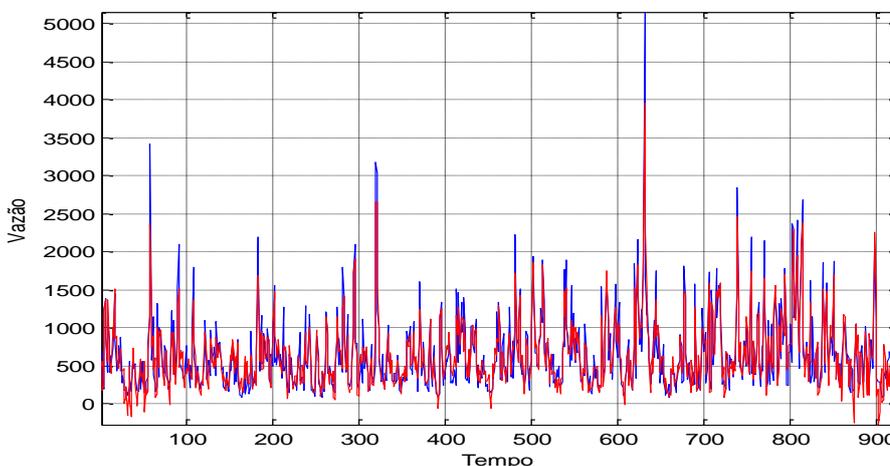
Na Figura 6, nota-se que os gráficos dos pares de vetores singulares (2,3), (10,11), e (14,15) mostram que existe, em cada par, associação *não linear* dos vetores singulares que o compõem, de modo que não são classificados como ruído.

Por outro lado, o gráfico de dispersão do par (9,10) mostra que estes dois vetores não estão associados. Dado que o vetor singular 10 não pode ser classificado como ruído (pois está associado ao vetor singular 11), existem evidências iniciais que o vetor singular 9 seja ruído (Hamilton, 1994). Por sua vez, o gráfico do par (2,3) evidencia que os vetores singulares 2 e 3 sejam componentes *harmônicas* passíveis de serem representadas por senos ou cossenos (Brock, Dechert, Scheinkman, 1987, Golyandina et al., 2001).

Em última análise, de acordo com (Hassani, 2007), pode-se verificar que os vetores singulares que compõem os pares (10,11) e (14,15) são componentes *harmônicas*, pois seus diagramas de dispersão exibem *polígonos regulares*. Em particular, tem-se que os vetores 10 e 11 possuem período igual a 7 meses, uma vez que o gráfico no plano cartesiano sugere a figura de um heptágono regular e os dados são mensais; enquanto que os vetores 14 e 15 possuem período igual a 4 meses, uma vez que o gráfico no plano cartesiano sugere a figura de paralelogramo regular.

Com base nestas interpretações, procedeu-se no agrupamento das componentes de tendência, harmônicas e de ruído com o uso do *software Caterpillar SSA* em 3 *clusters*. Com efeito, segue que 3 componentes SSA foram geradas e aquela classificada como ruído (via estatísticas BDS e de *Ljung-Box*), removida.

Na figura 7, tem-se a série temporal de Vazão da UHE GB Munhoz e a sua versão filtrada através da análise gráfica dos vetores singulares.



**Figura 7** - Série Temporal de Vazão e sua Aproximação através da Análise Gráfica de Vetores Singulares.

Na Tabela 1, têm-se os resultados do teste BDS e do teste de *Ljung-Box* aplicados sobre a série temporal de resíduos nos três casos estudados.

Na Tabela 2, têm-se os resultados do teste de *Dickey-Fuller* sobre a série temporal de resíduos e de correlação ponderada  $\rho_{1,2}^{(\omega)}$  entre a série temporal de resíduos oriundas das três abordagens (1) e a série temporal de vazão da UHE GB Munhoz filtrada (2).

**Tabela 1 - Testes BDS e de *Ljung-Box*.**

Método	Teste BDS				Teste de <i>Ljung-Box</i>				
	Dimensão	Estatística BDS	Estatística Z	Prob.	Lag	FAC	FACP	Estatística Q	Prob.
ACP	2	-0,0000297	-0,319459	0,7949	1	0,0038	0,0038	1,3043	0,253
	3	-0,0000422	-0,203157	0,8390	2	-0,0941	-0,0985	5,9841	0,104
	4	-0,0000139	-0,039952	0,9681	3	-0,0925	-0,0922	8,1544	0,095
	5	-0,0000624	-0,122574	0,9024	4	-0,0812	-0,0865	6,2540	0,134
	6	-0,0000228	-0,033133	0,9736	5	0,0027	0,0028	2,0687	0,352
Análise de Agrupamento integrada com ACP	2	-0,001258	-0,544596	0,5860	1	-0,024	-0,024	0,5517	0,458
	3	0,000671	0,183619	0,8543	2	-0,052	-0,053	3,0656	0,216
	4	0,002181	0,503658	0,6145	3	-0,004	-0,007	3,0813	0,379
	5	0,004454	0,991761	0,3213	4	0,005	0,002	3,1046	0,540
	6	0,005500	1,276060	0,2019	5	0,048	0,047	5,2170	0,390
Análise Gráfica dos Vetores Singulares	2	-0,0000412	-0,510289	0,5849	1	0,0037	0,0029	1,3210	0,228
	3	-0,0000917	-0,410241	0,6580	2	0,0908	0,0921	5,2365	0,204
	4	-0,0000566	-0,144583	0,9025	3	-0,0360	-0,0326	4,1524	0,423
	5	-0,0000244	-0,045879	0,9287	4	-0,0021	-0,0015	3,2860	0,564
	6	-0,0000302	-0,482100	0,6024	5	0,0031	0,0033	2,5897	0,583

**Tabela 2 - Testes de *Dickey-Fuller* e de Correlação Ponderada.**

Método	Teste de <i>Dickey Fuller</i>			Correlação Ponderada
	Estatística ADF	Nível	Valores Críticos	
ACP	-13,12375	1% 5% 10%	-3,4402 -2,8651 -2,5687	$\rho_{1,2}^{(\omega)} = 0,00123$
Análise de Agrupamento integrada com ACP	-12,96724	1% 5% 10%	-3,4402 -2,8651 -2,5687	$\rho_{1,2}^{(\omega)} = 0,00051$
Análise Gráfica dos Vetores	-13,22529	1% 5% 10%	-3,4402 -2,8651 -2,5687	$\rho_{1,2}^{(\omega)} = 0,00031$

A partir dos resultados apresentados na Tabela 1, pode-se concluir que as séries temporais de ruídos oriundas das três abordagens são estacionárias de segunda ordem. Adicionalmente, não são significativas as correlações ponderadas entre a série temporal de resíduos e a respectiva série temporal aproximada, conforme indicado nos resultados apresentados na Tabela 2.

## 6. Conclusões

Neste artigo, foi proposto o uso do método SSA para a filtragem de séries temporais. Para tanto, três abordagens foram propostas: ACP; ACP integrada com Análise de Agrupamento; e Análise Gráfica de Vetores Singulares. Para a ilustração dos métodos, foi utilizada a série temporal de vazão da UHE GB de Bento Munhoz.

Na separação entre as componentes de sinal e de ruído, foram utilizados testes estatísticos de *Ljung-Box*, BDS, correlação ponderada e de *Dickey-Fuller* e estes certificaram, nas Tabelas 1 e 2, que os resíduos removidos possuem propriedades estatísticas de ruído. Portanto, tem-se que as três abordagens propostas se mostraram eficientes no caso aplicado, não obstante são aplicáveis a qualquer série de tempo estocástica.

## Referências

- Aldefender, M.S. e Blashfield, R.K.** (1984), *Cluster Analysis*. Sage Publications. California. Pp. 35-45.
- Brock, W.A., W. Dechert, J. e Scheinkman,** (1987), A test for independence based on the correlation dimension. *Working paper*, University of Wisconsin at Madison, University of Houston, and University of Chicago.
- COPEL – Companhia Paranaense de Energia Elétrica,** (2012), Página da Internet: <http://www.copel.com/hpcopel/root/index.jsp>.
- Gámiz-Fortis, S. R., Esteban-Parra, M. J. e Castro-Diez, Y.** (2012), Advances in Streamflow Prediction: A Multimodel Statistical Approach for Application on Water Resources Management. Site: <http://www.intechopen.com/books/climate-models>.
- GISTATGROUP,** (2010), Caterpillar SSA. Petersburg University. Department of Mathematics. Russia. Site: <http://www.gistatgroup.com/cat/index.html>.
- Golyandina, N., Nekrutkin, V. e Zhihgljavsky, A.** (2001), *Analysis of time series structure: SSA and related techniques*. Chapman&Hall/CRC. New York, USA.
- Hamilton, J** (1994). *Time Series Analysis*. Princeton University Press.
- Hassani, H.** (2007). Singular Spectrum Analysis: Methodology and Comparison. *Journal of Data Science* 5, 239-257.
- Kubrusly, C. S.** (2001), *Elements of Operator Theory*. Birkhäuser, Boston.
- Morettin, P.A., Toloi, L.M.C.** (2006), *Análise Séries Temporais, 2ª Ed. ABE. Projeto Fisher*. Editora: Edgard Blucher.
- Vasconcelos, S.** (2011). Análise de Componentes Principais. Site: <http://www.ic.uff.br/~aconci/PCA-ACP.pdf>.
- Terry, L.A., Pereira, M.V.F., Silva, L.F. Sales, P.R., Araripe N. T.A.** (1986). Coordinating the Energy Generation of the Brazilian System. *Interfaces Special Issue*.