

## TÉCNICAS DE CLUSTERIZAÇÃO PARA O PROCESSO DE SELEÇÃO DAS AUTOTRIPLAS DO MODELO SSA NA PREVISÃO HORÁRIA DA VELOCIDADE DO VENTO

**Soraida Aguilar Vargas**

Pontifícia Universidade Católica do Rio de Janeiro - PUC-Rio  
saguilar@ele.puc-rio.br

**Reinaldo Castro Souza**

Pontifícia Universidade Católica do Rio de Janeiro - PUC-Rio  
reinaldo@ele.puc-rio.br

**José Francisco Pessanha**

Universidade do Estado do Rio de Janeiro - UERJ  
professorjfm@hotmail.com

### RESUMO

Neste trabalho a Análise Espectral Singular (SSA) é aplicada na previsão da média horária da velocidade do vento, uma informação importante para integração segura e econômica de parques eólicos ao sistema elétrico. A SSA decompõe uma série temporal em componentes que podem ser classificadas em três categorias: tendência, componentes harmônicas e ruídos. Na sequência, as componentes ruidosas são descartadas e ao final obtém-se o sinal, a informação relevante para realização da previsão. Tradicionalmente a classificação das componentes é efetuada por meio de análises pictóricas de resultados da SSA, contudo a aplicação da SSA na modelagem e previsão de séries horárias tende a produzir um elevado número de componentes dificultando a inspeção visual. Neste trabalho é investigado o uso da análise de agrupamentos na classificação das componentes resultantes da aplicação da SSA na previsão da média horária da velocidade do vento.

**PALAVRAS-CHAVE.** Velocidade do vento, análise espectral singular, clusterização.

### ABSTRACT

In this work the Singular Spectrum Analysis (SSA) is applied to the forecasting of mean hourly wind speed time series, an important information for safe and economical integration of wind farms to the electrical system. The SSA decomposes a time series into components that can be classified in three categories: trend, harmonic components and noise. As a result, the noisy components are discarded resulting, in the end, the signal, the contain the relevant information to perform the prediction. Traditionally, the classification of the components is effected using pictorial analysis results of SSA, however the implementation of the SSA in the modeling and prediction of hourly time series tends to produce a huge number of components that make difficult the visual inspection. This work investigates the use of cluster analysis in classification of the components resulting from the application of SSA in predicting the hourly mean wind speed.

**KEYWORDS.** Wind speed, singular spectrum analysis, clustering.

## 1. Introdução

O Plano Decenal de Expansão de Energia – PDE 2023 (Brasil, 2014) indica que a participação da energia eólica na matriz de geração elétrica deverá alcançar 8,1% em 2023, em função da ampliação de cerca de 20 GW da capacidade instalada em parques eólicos no mesmo período. A maior participação da geração eólica contribui para a diversificação da matriz de geração elétrica, complementando a geração predominantemente hidroelétrica, sobretudo nos períodos de seca (Palfi & Zambon, 2013).

Contudo, a geração eólica depende da velocidade do vento, um recurso renovável e abundante, porém uma variável aleatória. Em função do comportamento estocástico do vento, a integração de parques eólicos na operação de sistemas elétricos baseia-se em previsões da velocidade do vento em horizontes de curto prazo, por exemplo, até 24 horas à frente em bases horárias.

A previsão de curto prazo da velocidade do vento pode ser obtida por diferentes métodos para previsão de séries temporais (Zhu e Genton, 2012), por exemplo, os métodos estatísticos (ARIMA, Filtro de Kalman, previsão probabilística, etc.), métodos de inteligência computacional (redes neurais artificiais, máquinas de vetor de suporte, lógica *fuzzy* e sistemas *neuro-fuzzy*) e métodos híbridos que combinam abordagens baseadas em técnicas estatísticas e inteligência computacional.

Recentemente, a análise espectral singular (SSA – *Singular Spectral Analysis*) (Golyandina et al., 2001; Hassani, 2007) tem emergido como mais uma técnica para previsão da velocidade do vento. Basicamente, a SSA decompõe uma série temporal em um grande número de componentes que podem ser classificadas em três categorias: tendência, componentes harmônicas e ruídos. Na sequência, as componentes ruidosas são descartadas e ao final obtém-se o sinal, ou seja, a informação relevante e útil para realização da previsão.

Em geral a SSA tem sido aplicada em séries mensais, nas quais o número de componentes tende a ser pequeno e a classificação das componentes pode ser efetuada por meio de análises pictóricas. Contudo, a aplicação da SSA em séries horárias tende a produzir um grande número de componentes a serem classificadas o que torna inviável a classificação por meio de inspeções visuais dos resultados obtidos. Neste caso a classificação deve ser efetuada por meio de métodos automáticos para análise de agrupamentos (*cluster analysis*). Esta dificuldade decorre do fato da modelagem da estrutura de autocorrelação presente em uma série horária demandar um grande número de observações.

Portanto, o objetivo deste trabalho consiste em investigar a classificação das componentes por meio de métodos de *cluster analysis* aglomerativos com a finalidade de melhorar a performance da SSA na previsão da média horária da velocidade do vento. O trabalho está organizado em cinco seções. A seguir, na seção 2 apresenta-se uma breve descrição da metodologia SSA. Em seguida na seção 3 tem-se uma breve introdução aos métodos hierárquicos aglomerativos. Na seção 4, descrevem-se os resultados de um experimento computacional realizado com uma série de velocidade de vento. Por fim, na seção 5 são apresentadas as principais conclusões do trabalho.

## 2. Análise Espectral Singular (SSA)

A SSA é uma técnica não paramétrica para análise e previsão de séries temporais e, portanto, baseia-se exclusivamente nos dados (Beneki et al., 2009). Adicionalmente, a SSA dispensa a premissa de estacionariedade da série temporal (Hassani, 2007).

A SSA decompõe a série de interesse em várias componentes aditivas. Contudo, de forma distinta da abordagem de Fourier que emprega em funções trigonométricas na decomposição de uma série temporal, a SSA produz um conjunto de componentes diretamente da série temporal em estudo sem se basear em funções específicas.

Basicamente, a aplicação da SSA envolve três etapas: decomposição, reconstrução e previsão. No primeiro estágio, decompõe-se a série em diversas componentes classificadas em

tendência, harmônicas e ruído. No segundo estágio, a série é reconstruída sem as componentes do ruído e assim utilizada no terceiro e último estágio para a previsão. Os três estágios supracitados são descritos nas seções a seguir.

## 2.1. Decomposição

A etapa de decomposição tem duas fases: incorporação e decomposição em valores singulares. Na fase de incorporação a série temporal de interesse é representada como uma sequência de vetores de comprimento  $L$  que formam as colunas da matriz trajetória. Na sequência faz-se a decomposição em valores singulares (SVD – *Singular Value Decomposition*) da matriz de trajetória.

### 2.1.1. Incorporação

Na fase de incorporação uma janela móvel de tamanho  $L$  percorre a série temporal  $Y_T = (y_1, y_2, \dots, y_T)$  e a cada passo da janela obtém-se uma coluna da matriz trajetória em (1), cujo total de colunas é igual a  $K=T-L+1$ .

$$\mathbf{X} = \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_K \\ y_2 & y_3 & y_4 & \cdots & y_{K+1} \\ y_3 & y_4 & y_5 & \cdots & y_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \cdots & y_T \end{bmatrix} \quad (1)$$

O tamanho da janela móvel  $L$  determina o número de componentes extraídas da série tempo.  $L$  deve assumir um valor inteiro suficientemente grande, mas não superior a  $T/2$  (Golyandina et al., 2001; Hassani, 2007).

A matriz de trajetória  $\mathbf{X}$  em (1) possui uma estrutura semelhante à de uma matriz de Hankel, o que significa que todos os elementos  $\{a_{ij}\}$  ao longo das diagonais paralelas à diagonal secundária são iguais, além de ser uma matriz simétrica. Note-se também que na matriz trajetória a primeira coluna  $(y_1, \dots, y_L)$  e a última linha  $(y_L, \dots, y_T)$  formam a série original.

### 2.1.2. Decomposição em Valores Singulares (SVD)

Por meio da SVD a matriz de trajetória  $\mathbf{X}$  em (1) pode ser expressa pela seguinte soma de matrizes elementares:

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_d \quad (2)$$

em que  $d$  é o número de autovalores ( $\lambda_1 \geq \dots \lambda_L \geq 0$ ) não nulos da matriz  $\mathbf{S} = \mathbf{X}\mathbf{X}^T$  e  $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$ , sendo que  $(U_1, \dots, U_L)$  são os autovetores de  $\mathbf{S} = \mathbf{X}\mathbf{X}^T$  e  $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i} \forall i = 1, \dots, d$ .

Uma autotripla corresponde ao conjunto  $\{U_i, V_i, \lambda_i\} \forall i = 1, \dots, d$ . A informação contida em cada autotripla é quantificada pela razão entre os autovalores e seu somatório  $\lambda_i / \sum_{i=1}^d \lambda_i$ .

## 2.2. Reconstrução

A reconstrução ocorrer em duas fases: *agrupamento* e *média diagonal*. Na fase de *agrupamento* as  $d$  matrizes elementares obtidas na fase de SVD são agrupadas em grupos mutuamente exclusivos (*clusters*). Na sequência, as matrizes são convertidas em séries temporais

por meio do procedimento denominado média diagonal e ao final são obtidas as componentes aditivas que formam a série temporal.

### 2.2.1. Agrupamento das Autotriplas

Nesta fase, o conjunto de índices  $\{1, \dots, d\}$  é segmentado em  $m$  ( $m < d$ ) subconjuntos disjuntos  $\{I_1, \dots, I_m\}$ , de tal forma que as matrizes elementares  $\mathbf{X}_i$  correspondentes ao índices em um mesmo grupo sejam classificadas em um mesmo *cluster* e na sequência somadas. Assim, a matriz trajetória pode ser expressa pela soma de  $m$  matrizes.

$$\mathbf{X} = \mathbf{X}_{I_1} + \mathbf{X}_{I_2} + \dots + \mathbf{X}_{I_m} \quad (3)$$

Por exemplo, se  $d=6$  e  $m=3$ , as seis matrizes elementares são agrupadas em três *clusters*, conforme ilustrado na Figura 1.

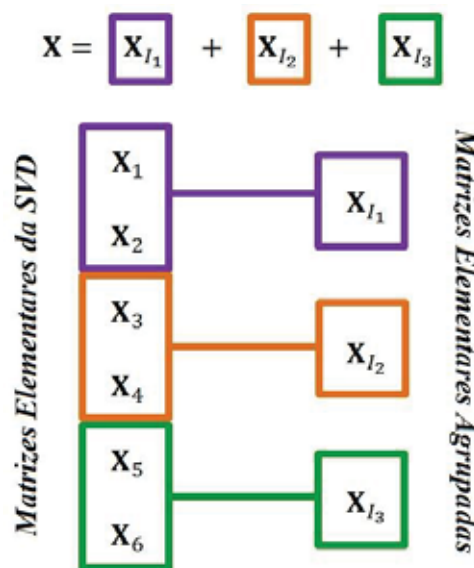


Fig. 1. Agrupamento das matrizes elementares.

### 2.2.2. Média Diagonal Secundaria

Nesta fase, cada matriz resultante do processo de agrupamento é transformada em uma série temporal de tamanho  $T$  (Golyandina & Zhigljavsky, 2013). Considere uma matriz  $\mathbf{X}_{I_s} \forall s=1, m$  com elementos  $x_{ij}, 1 \leq i \leq L, 1 \leq j \leq K$ . Seja  $L^* = \min(L, K)$ ,  $K^* = \max(L, K)$  e  $T = L + K - 1$ . Adicionalmente, seja  $x_{ij}^* = x_{ij}$ , se  $L < K$  e  $x_{ij}^* = x_{ji}$ , caso contrário. Assim, as fórmulas a seguir transformam a matriz  $\mathbf{X}_{I_s}$  na série  $\tilde{x}_1, \dots, \tilde{x}_T$  da seguinte maneira:

$$\tilde{x}_k^{(s)} = \begin{cases} \frac{1}{k} \sum_{m=1}^k x_{m, k-m+1}^* & \text{para } 1 \leq k < L^*, \\ \frac{1}{L^*} \sum_{m=1}^{L^*} x_{m, k-m+1}^* & \text{para } L^* \leq k \leq K^*, \\ \frac{1}{T-k+1} \sum_{m=k-K^*+1}^{T-K^*+1} x_{m, k-m+1}^* & \text{para } K^* < k \leq T. \end{cases} \quad (4)$$

O procedimento descrito em (4) corresponde a uma média dos elementos da matriz sobre diagonais paralelas à diagonal secundária tal que  $i + j = k + 1$ , com  $k = 1, 2, \dots, T$ . As médias diagonais da matriz  $\mathbf{X}_{I_s} \forall s=1, \dots, m$  formam a série  $\tilde{X}^{(s)} = (\tilde{x}_1^{(s)}, \dots, \tilde{x}_T^{(s)})$ . Portanto, a série inicial  $y_1, \dots, y_T$  é decomposta em uma soma de  $m$  séries:

$$y_t = \sum_{j=1}^m \tilde{x}_t^{(j)} \quad \forall t = 1, \dots, T \quad (5)$$

Removendo as componentes irregulares (ruídos) obtém-se uma versão suavizada da série temporal, ou seja, uma estimativa do sinal  $\hat{x}$  contido na série temporal. Por exemplo, admitindo que o sinal seja formado pelas  $r$  primeiras componentes, o sinal  $\hat{x}$  é definido pela soma das  $r$  primeiras séries, enquanto o ruído  $\hat{\varepsilon}$  corresponde ao total das demais componentes. Assim, uma série temporal  $x$  pode ser decomposta em duas parcelas aditivas  $\hat{x} + \hat{\varepsilon}$  onde:

$$\hat{y}_n = \sum_{j=1}^r \tilde{x}_n^{(j)} \quad \forall n = 1, \dots, T \quad (6)$$

$$\hat{\varepsilon}_n = \sum_{j=r+1}^m \tilde{x}_n^{(j)} \quad \forall n = 1, \dots, T \quad (7)$$

### 2.3. Previsão

Para obter previsões via SSA o requisito básico é que a série temporal satisfaça as relações de recorrência linear (LRF - Linear Recurrent Formula), conforme indicado em Golyandina e Zhigljavsky (2013). Uma série  $Y_T = (y_1, y_2, \dots, y_T)$  satisfaz a LRF de ordem  $L-1$  se:

$$y_i = \sum_{j=1}^{L-1} a_j y_{i-j}; \quad i = T + 1, \dots, T + M \quad (8)$$

O principal pressuposto é que o comprimento da janela  $L$  foi escolhido de forma a separar o sinal do ruído. As previsões são realizadas utilizando-se as  $r$  autotriplas escolhidas, conforme indicado a seguir:

$$\hat{y}_i = \begin{cases} \tilde{X}_i, & i = 1, \dots, T \\ \sum_{j=1}^{L-1} a_j \hat{y}_{i-j}, & i = T + 1, \dots, T + M \end{cases} \quad (9)$$

em que  $\tilde{X}_1, \dots, \tilde{X}_T$  denotam os valores da série reconstruída obtidos em (4), enquanto que  $\hat{y}_{T+1}, \dots, \hat{y}_{T+M}$  indicam as previsões  $M$  passos à frente.

Ainda em (9) os coeficientes  $a_j$  formam o vetor  $R = (a_1, \dots, a_{L-1})^T$  determinado como segue (Mahmoudvand et al, 2013):

$$R = \frac{1}{1 - v^2} \sum_{j=1}^r \pi_j U_j^v, \quad (10)$$

em que o vetor  $U_j^v$  denota as primeiras  $L - 1$  componentes do autovetor  $U_j$ ,  $\pi_j$  é o último elemento de  $U_j$  ( $j = 1, \dots, r$ ) e  $v^2 = \sum_{j=1}^r \pi_j^2$ .

### 3. Métodos Hierárquicos de Agrupamento

Na fase de agrupamento do estágio de reconstrução são empregadas diferentes técnicas de análise de agrupamento para fazer a agregação das matrizes elementares com maior similaridade. Entre os métodos mais utilizados em SSA para este propósito encontram-se *Análise dos Valores Singulares*, a *Análise Gráfica de Vetores Singulares* e a *Matriz de Correlações Ponderadas*, auxiliado pelo *periodograma* dos autovalores. No entanto, uma problemática latente é que todas estas técnicas envolvem análises pictóricas e devem ser usadas simultaneamente para poder fazer uma melhor escolha dos conjuntos de componentes mais semelhantes. O problema é agravado pelo fato de que a quantidade de componentes cresce com o aumento do tamanho da janela  $L$ , em geral fixado em metade da quantidade de observações da série temporal. Assim, dependendo do tamanho da série a classificação por meio de análises pictóricas das componentes pode se tornar inviável. Para superar esta dificuldade são utilizados os métodos de análise de agrupamentos, em especial métodos hierárquicos aglomerativos.

Os métodos hierárquicos são classificados entre aglomerativos e divisivos, sendo que os primeiros são os mais amplamente utilizados (Upton & Cook, 2008), pois em função do maior grau de complexidade computacional dos métodos divisivos (Reis, 2001) os métodos aglomerativos são facilmente encontrados em diversos *softwares*.

O algoritmo hierárquico aglomerativo inicia-se considerando cada indivíduo como sendo um *cluster*, totalizando  $n$  *clusters*. Em cada etapa, calcula-se a distância entre cada par de *clusters*. Estas distâncias são geralmente, armazenadas em uma matriz de dissimilaridade simétrica. Posteriormente, escolhem-se dois *clusters* com a distância mínima e logo são unidos. A seguir, atualiza-se a matriz de distâncias. Este processo continua iterativamente até que todos os indivíduos estejam em um único *cluster* (o nível mais alto da hierarquia) (Linden, 2009) ou até que uma condição de término ocorra (Agrawal et al., 1998; Ng & Han, 1994; Han & Kamber, 2001).

Para o cálculo das *medidas de similaridade* podem ser utilizadas métricas tais como: distância euclidiana, *Distância de Minkowski*, *distância de Manhattan*, *distância de Chebychev*, *Distância de Canberra*, *Coefficiente de Correlação de Pearson* ou o *Coefficiente de Separação Angular*, sendo esta última a métrica escolhida para o cálculo da medida de similaridade da matriz de similaridades, a qual está dada pela seguinte expressão:

$$c_{ij} = \frac{\sum_{k=1}^p x_{ik} \cdot x_{jk}}{\left(\sum_{k=1}^p x_{ik}^2 \sum_{k=1}^p x_{jk}^2\right)^{1/2}} \quad (11)$$

Os resultados do método hierárquico são exibidos de forma gráfica através de uma árvore denominada dendrograma (Figura 2), a qual representa as sucessivas fusões dos *clusters*. Localizada no topo do dendrograma, a raiz representa o agrupamento de todos os indivíduos, enquanto os nós na base do dendrograma (folhas) representam os indivíduos. O resultado do agrupamento pode ser obtido ao cortar o dendrograma em diferentes níveis de acordo com o número de *clusters*  $K$  desejados. Esta forma de representação fornece descrições informativas e uma visualização para as estruturas de grupos potenciais. Em tais hierarquias, cada nó da árvore representa um *cluster* da base de dados.

Fazendo uso do dendrograma e com base no conhecimento sobre a estrutura dos dados, deve-se estabelecer uma distância de corte para determinar quais serão os grupos formados. Essa decisão é subjetiva, e deve ser feita de acordo o objetivo da análise e o número de grupos desejados.

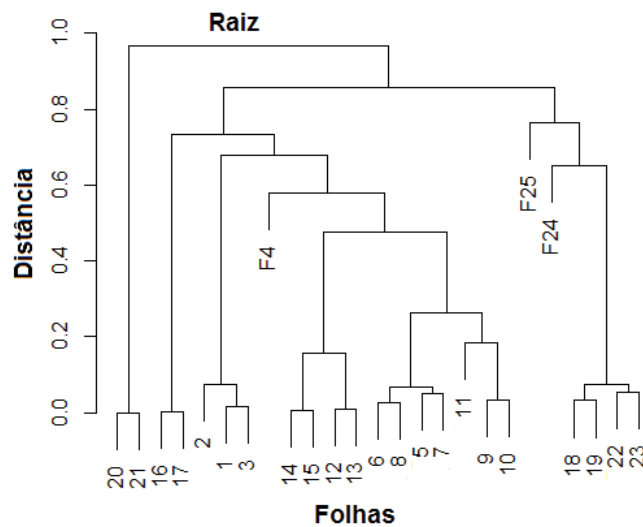


Fig. 2. Dendrograma.

#### 4. Experimento computacional

Para ilustrar a aplicação dos métodos descritos acima considerou-se uma série temporal de médias horárias da velocidade do vento em uma determinada localidade ao longo do período entre 1 de janeiro de 2007 até 1 de Janeiro de 2008, perfazendo um total de 8.784 observações. Na modelagem da série foram consideradas as primeiras 8760 observações, enquanto as 24 restantes foram usadas para fins de validação (*out-of-sample*). Na Figura 3 apresenta-se a série temporal das médias horárias da velocidade do vento, caracterizada por padrões sazonais diário e mensal.

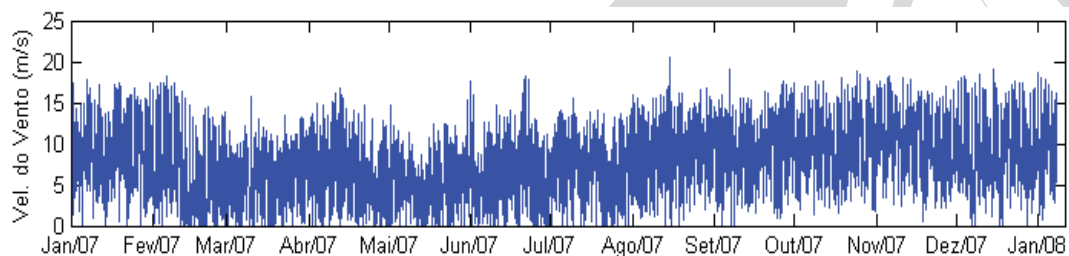


Fig. 3 – Série temporal da velocidade do vento (m/s).

##### 4.1. Identificação dos valores de $L$ e $r$ por enumeração

Para o hiperparâmetro  $L$  foram avaliados múltiplos da sazonalidade horária, conforme indicado na primeira coluna da Tabela 1. No caso do hiperparâmetro  $r$ , a escolha foi feita por enumeração, onde foram avaliados de forma incremental os distintos valores até encontrar o valor que minimiza o erro quadrático médio (RMSE) entre a série observada e a série prevista. Desta forma os hiperparâmetros  $L$  e  $r$  foram fixados em 720 e 18 respectivamente, conforme é indicado na Tabela 1.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_t - y_t)^2}{N}} \quad (12)$$

Tabela 1. Parâmetro de defasagem  $L$ .

Parâmetro $L$	Parâmetro $r$	RMSE
24	1 até 10	1,9557
36	1 até 16	2,0143
168	1 até 25	1,9217
720	1 até 18	<b>1,4822</b>
2920	1 até 28	2,6318
4380	1 até 29	2,4642

#### 4.2. Identificação dos valores $r$ por análise de agrupamento hierárquico.

Com a identificação de  $L = 720$  na seção anterior, inicializa-se a segunda fase do estágio de decomposição de SSA, para posteriormente efetuar o agrupamento das autotriplas por meio da análise de agrupamentos, em particular pelo método hierárquico aglomerativo, como foi indicado anteriormente. O processo requer da escolhida de uma medida de similaridade, neste caso, foi adotada a matriz de correlações ponderadas como medida de similaridade pela sua semelhança com o “*Coeficiente de separação angular*” (Eq. 11), A matriz de correlações ponderadas foi calculada pelo pacote Rssa disponível no software “**R Project for Statistical Computing**”, em que cada elemento é uma medida da associação linear entre duas séries  $Y_T^{(1)}$  e  $Y_T^{(2)}$ :

$$\rho_{12}^{(\omega)} = \frac{(Y_T^{(1)}, Y_T^{(2)})_{\omega}}{\|Y_T^{(1)}\|_{\omega} \|Y_T^{(2)}\|_{\omega}} \quad (13)$$

Os elementos da matriz de de correlações ponderadas ( $s_{ij}$ ) pode ser tranformada em uma matriz de dissimilaridade ( $d_{ij}$ ) por meio da seguinte a operação  $d_{ij} = 1 - s_{ij}$ . Adicionalmente é preciso estabelecer o critério de agregação de *clusters*, o qual foi escolhido como sendo o Método do vizinho mais próximo ou *single-link*. A especificação de todos estes procedimentos fornece os dendrogramas para as autotriplas da SVD. Para uma melhor visualização dos *clusters* foram elaborados dendrogramas com matrizes de *correlação ponderada* compostas por 50 autotriplas aproximadamente, isto devido ao fato de que o total de autotriplas é da ordem de 720 impossibilitando a inspeção visual. A Figura 4 mostra o dendrograma para as primeiras 50 autotriplas da SVD.



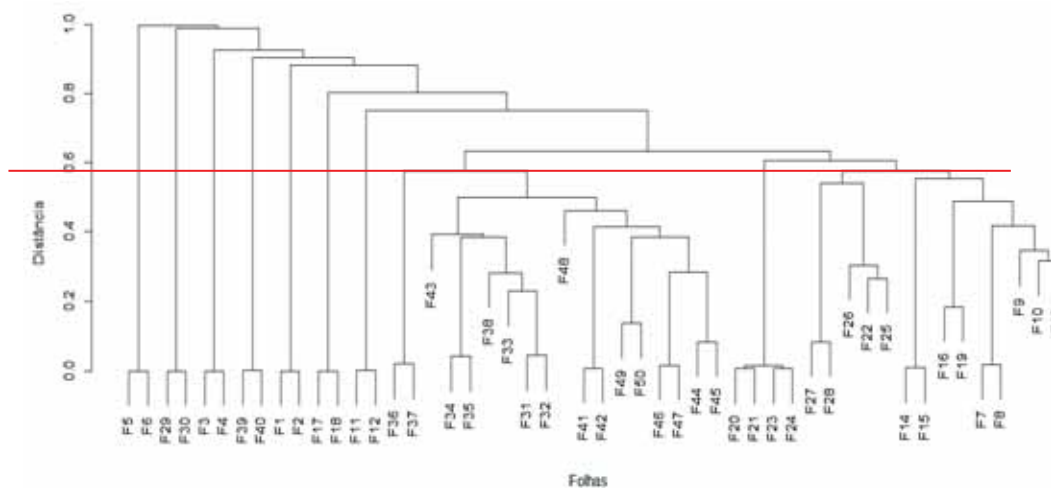


Fig. 4. Dendrograma das autotriplas 1 até 50.

De forma análoga foram calculados os dendrogramas para as restantes 670 autotriplas, com os quais se determinou os grupos de autotriplas para a reconstrução da série temporal (Tabela 2), permitindo com isto fazer o cálculo das previsões 24 horas à frente como se mostra na Figura 5.

Tabela 2. Autotriplas escolhidas pelo Método Hierárquico

D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14		
1	54	101	151	201	251	308	351	404	456	501	564	604	651	677	700
2	55	102	152	202	252	309	352	405	457	502	565	605	652	678	701
3	63	103	153	203	253	312	357	426	459	506	582	649	653	679	702
4	64	104	154	204	266	313	358	427	460	507	583	650	654	680	707
5	67	113	155	205	267	314	359	445	461	520	598		655	681	708
6	68	114	156	206	268	315	360	446	462	521	599		656	682	709
11	71	117	157	207	269	316	361		464	524			657	683	710
12	72	118	158	208	270	317	362		465	525			658	684	711
17	78	136	168	209	271	318	363		466	528			659	685	712
18	79	137	169	218	272	319	364		467	529			660	686	713
20	88	138	174	219	273	320	365		468	547			664	687	714
21	89	139	175	220	286	321	366		469	548			665	694	715
23	90		176	221	287	322	367		475	549			668	695	716
24	91		177	222	297	324	368		476	550			669	696	717
25	97		178	223	298	325	380		477				672	697	718
26	98		179	224	299	326	381		478				673	699	719
29			196	229	300	327	383		491				674		720
30			197	230		328	384		492						
39				231		329	386		493						
40				232		330	387		494						
				233			388		495						
				234			389		496						
				235			390		497						
				236			391		498						
				241					499						
				242					500						

Pode-se observar que com o uso de análise de agrupamento o valor do RMSE quando o  $r$  é determinado por enumeração (1,4822) diminui se comparado com valor de  $r$  calculado através de análise de agrupamento hierárquico (1,2050), dado que existem componentes que contem informação que não está contida no processo tradicional.

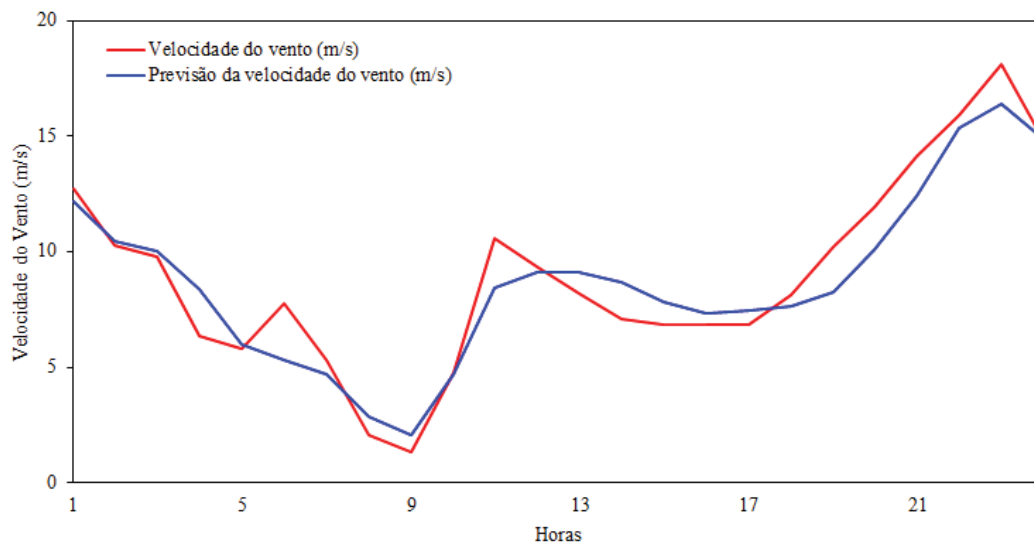


Figura 5 – Previsão 24 horas à frente da Velocidade do Vento usando a metodologia SSA.

Com o intuito de avaliar os resultados obtidos com SSA foram implementados modelos que envolvem um enfoque diferente de previsão da velocidade do vento, por exemplo, os modelos SARIMA (Box & Jenkins, 1970), os modelos *Double Seasonal Holt-Winters* (Taylor, 2003), além do tradicional método de persistência conhecido também como previsão ingênua. Os resultados podem ser observados na Figura 6. Adicionalmente as medidas de erro para cada um dos modelos são apresentadas na Tabela 3, na qual destaca-se a SSA com as melhores previsões.

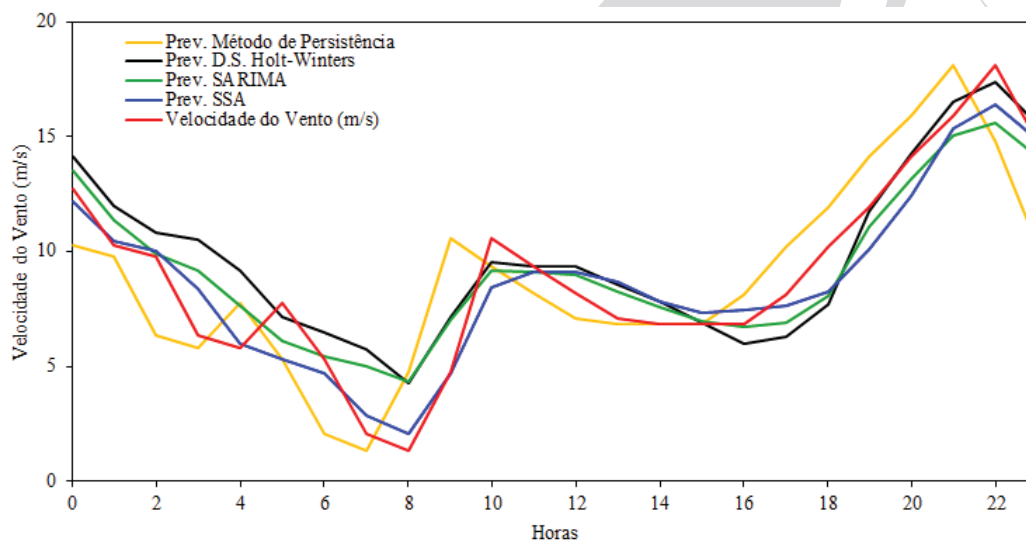


Fig. 6 – Previsão 24 horas à frente da Velocidade do Vento para diferentes modelos.

Tabela 3. Medidas de desempenho das abordagens avaliadas.

	SSA	SARIMA	D.S. Holt-Winters	Persistência
<b>RMSE (m/s)</b>	<u>1,2050</u>	1,5630	1,8467	2,4377
<b>MAD (m/s)</b>	<u>0,9585</u>	1,2692	1,4500	1,9576
<b>U-THEIL</b>	<u>0,5345</u>	0,6932	0,8191	1,0008

Este resultado é muito importante porque ele faz parte do processo de previsão da geração de energia eólica, que na maioria das vezes, é feito em dois estágios; onde no primeiro estágio se calcula a previsão da velocidade do vento para diferentes horizontes de tempo e no segundo estágio é utilizada a curva de potência e a velocidade do vento prevista para fornecer assim a previsão da produção eólica. Cabe anotar que no caso de não possuir previsões o suficientemente precisas, os erros na previsão da velocidade são amplificados na previsão da geração de energia eólica, visto que a potência gerada é proporcional ao cubo da velocidade do vento.

## 6. Conclusões

A aplicação da Análise Espectral Singular (SSA) na previsão da média horária da velocidade do vento se mostrou efetiva e apresentou um desempenho compatível com as abordagens tradicionais. A SSA decompõe uma série temporal em componentes que podem ser classificadas em três categorias: tendência, componentes harmônicas e ruídos. Tradicionalmente a classificação das componentes é efetuada por meio de análises pictóricas de resultados da SSA, contudo a aplicação da SSA na modelagem e previsão de séries horárias tende a produzir um elevado número de componentes comprometendo a eficiência da inspeção visual. A proposta de aplicação da análise de agrupamentos aglomerativa *single-linkage* na classificação das componentes oriundas da SSA permitindo a extração eficiente do sinal contido na série das médias horárias da velocidade do vento. Como trabalho futuro, recomenda-se pesquisar novas técnicas de análise de agrupamentos, bem como a abordagem MSSA com o objetivo de levar em consideração outras variáveis como a direção do vento, a temperatura que recolham maior informação da dinâmica de funcionamento do fenômeno.

## Agradecimentos

Os autores gostariam de agradecer à FAPERJ/CAPES pelo apoio financeiro.

## References

- Agrawal, R., Gehrke, J., Gunopulos, D., & RAGHAVAN, P. (1998). Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications. *IBM Almaden Research Center*.
- Beneki, C., Eeckels, B., & Leon, C. (2009). Signal extraction and forecasting of the UK tourism income time series. A singular spectrum analysis approach, MPRA paper no. 18354. Online at [http://mpr.aub.uni-muenchen.de/18354/1/MPRA\\_paper\\_18354.pdf](http://mpr.aub.uni-muenchen.de/18354/1/MPRA_paper_18354.pdf)
- Box, G. E. E. e Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, Cal.
- Brasil, Ministério das Minas e Energia, Empresa de Pesquisa Energética Plano Nacional de Expansão de Energia, Brasília: MME/EPE, 2014.

- Golyandina, N., Nekrutkin, V., & Zhigljavsky, A. (2001). *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman & Hall/CRC, New York.
- Golyandina, N., & Zhigljavsky, A. (2013). *Singular Spectrum Analysis for Time Series*. New York: Springer.
- Han, J., & Kamber, M. (2001). *Cluster Analysis*. In: Morgan Publishers (eds.), *Data Mining: Concepts and Techniques*, 1 ed., chapter 8, New York, USA, Academic Press.
- Hassani, H. (2007). Singular spectrum analysis: methodology and comparison. *Journal of Data Science*, 5, 239–257.
- Linden, R. (2009). Técnicas de Agrupamento. *Revista de Sistemas de Informação da FSMA*, 4, 18 – 36.
- Mahmoudvand, R., Najari, N. & Zokaei, M. (2013). On the optimal parameters for reconstruction and forecasting in the singular spectrum analysis. *Communication in Statistics - Simulations and Computations*, 42, 860-870.
- Ng, R. T., & Han, J. (1994). Efficient and Effective Clustering Methods for Spatial Data Mining. In *Proceedings of the 20th VLDB Conference*, Santiago-Chile, (pp. 144 – 155).
- PALFI, G.; ZAMBON, R. Hydro and Wind Power Complementarity and Scenarization in Brazil. World Environmental and Water Resources Congress, Cincinnati, 2013.
- Taylor, J. W. (2003). Short-Term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing. *Journal of Operational Research Society*, 54, 799–805.
- Zhu, X., e Genton, M. G. (2012). Short-Term Wind Speed Forecasting for Power System Operations. *International Statistical Review*, 80(1), 2–23.