

SINGULAR SPECTRUM ANALYSIS: ESTUDOS COMPARATIVOS DE PREVISÃO EM SÉRIES TEMPORAIS

Renata de Miranda Esquivel

Centro Integrado de Manufatura e Tecnologia (CIMATEC)/ SENAI
Av. Orlando Gomes, 1845- Piatã. Salvador-BA
esquivel.renata@hotmail.com

Valter de Senna

Centro Integrado de Manufatura e Tecnologia (CIMATEC)/ SENAI
Av. Orlando Gomes, 1845- Piatã. Salvador-BA
valter.senna@gmail.com

Gecynalda Soares da Silva Gomes

Departamento de Estatística da UFBA
Av. Adhemar de Barros, s/n- Campus de Ondina. Salvador-BA
gecynalda@yahoo.com

RESUMO

Os métodos estatísticos para análise de séries temporais encontram-se amplamente desenvolvidos na literatura e vários modelos clássicos preditivos estão disponíveis em *softwares* estatísticos. Contudo, cada modelo clássico exige suposições relacionadas às características dos dados e o uso adequado dos modelos exigirá verificações dessas suposições, o que pode demandar esforços na etapa de identificação do padrão de comportamento da série temporal. Como alternativa, pode-se utilizar a técnica conhecida como *Singular Spectrum Analysis* (SSA). A SSA realiza uma decomposição da série temporal em poucos componentes independentes. Este método não exige o conhecimento sobre o modelo paramétrico da série temporal e pode ser aplicado em qualquer série com alguma estrutura potencial. O presente artigo objetiva avaliar a capacidade preditiva da SSA comparando-a com alguns modelos clássicos para séries temporais. Com esta finalidade, examinamos duas séries temporais com características distintas: uma série proveniente da área da meteorologia e uma série gerada artificialmente.

PALAVRAS CHAVE. *Singular Spectrum Analysis*, Séries temporais, Previsão.

Área principal (Estatística)

ABSTRACT

Statistical methods for the analysis of time series abound in the literature and classical predictive models are implemented in many statistical *softwares*. However, each classical model requires assumptions related to the characteristics of the data and appropriate use of the models requires verification of these assumptions. Alternatively, one can use a powerful technique known as Singular Spectral Analysis (SSA). The SSA aims at a decomposition of the series in a few independent components. This method does not require knowledge of the parametric model of the series and can be applied in series with any potential structure. This article aims to evaluate the predictive ability of the SSA by comparing it with some classical models for time series. To this end we examined two series with different features, one containing meteorological data and the other an artificially generated series.

KEYWORDS. *Singular Spectrum Analysis*. Time series. Forecast.

Main area (Statistics)

1. Introdução

O desenvolvimento de métodos estatísticos para análise de dados obtidos em situações em que as observações são dependentes tem apresentado crescimento vertiginoso nas últimas décadas e, em particular, a análise de dados provenientes de séries temporais.

Ao trabalhar com séries temporais o objetivo mais usual é a predição de valores futuros. A necessidade de obter previsões precisas de eventos futuros ou suas consequências, sejam climáticas, econômicas, epidemiológicas ou de qualquer natureza, tem levado a um constante desenvolvimento de técnicas de previsão em séries temporais.

Os métodos estatísticos clássicos para análise de séries temporais encontram-se bem documentados na literatura pertinente. Contudo, boa parte desses métodos requer um conhecimento especializado para sua correta aplicação. Sendo assim, o uso adequado dos modelos clássicos exigirá verificações das suas suposições, o que demanda esforços e experiência, na análise exploratória dos dados. A Análise Espectral Singular (SSA, do inglês *Singular Spectrum Analysis*), se apresenta como uma alternativa relativamente simples e poderosa.

A SSA é um método não paramétrico usado na análise de séries temporais e que exige pouco conhecimento prévio do comportamento da série. Essa técnica investiga o comportamento das séries históricas através de uma decomposição e reconstrução dos seus componentes constitutivos, caracterizando os estágios da SSA.

A ferramenta SSA, na literatura técnica, tem-se mostrado útil nas análises de séries das áreas de meteorologia, geofísica, física, climatologia, economia, saúde e em vários outros campos do conhecimento. Essa ferramenta pode ser aplicada em séries curtas ou longas, séries não estacionárias ou estacionárias, ruidosas ou não, ou seja, em qualquer série temporal com alguma estrutura (Golyandina, Nekrutkin e Zhigljavsky, 2001; Hassani, 2007).

Vale salientar que, a SSA não se detém a realizar previsões de valores futuros, mas a técnica também tem a finalidade de identificar e extrair padrões geradores da série temporal.

O presente artigo objetiva realizar um estudo comparativo entre o algoritmo de previsão SSA e algumas estratégias preditivas clássicas. Com esta finalidade, analisamos duas séries temporais com características distintas: uma série proveniente da área da meteorologia (estudo empírico) e uma série artificial (estudo simulado).

2. Materiais e Métodos

2.1 Técnicas clássicas

Os métodos clássicos - que comumente apresentam bom poder preditivo - utilizados nesse trabalho comparativo, são os algoritmos de alisamento exponencial de Holt (SEH) e de Holt-Winters (H-W), e os métodos de Box e Jenkins (os modelos autorregressivos integrados de médias móveis (ARIMA) e a classe de modelos autorregressivos integrados de médias móveis sazonais (SARIMA)).

A descrição dessas estratégias preditivas clássicas pode ser encontrada em muitos referenciais teóricos, e.g. Box e Jenkins (1970); Brillinger (2001); Brockwell e Davis (2002); Morettin e Toloi (2006) etc.

O uso adequado do algoritmo de Holt-Winters se faz em decorrência da presença de sazonalidade nos dados da série, independentemente da presença da tendência. Já o algoritmo de Holt é adequado se existe basicamente uma tendência linear na série. Para a utilização do método de Box e Jenkins se exige conhecimento sobre as propriedades das classes de modelos, além da experiência do analista, no que tange a identificação do modelo mais apropriado aos dados. Fatos que dificultam a aplicação da metodologia de Box e Jenkins.

2.2 Singular Spectrum Analysis (SSA)

2.2.1 Breve descrição do SSA Básico

A técnica SSA básica fundamenta-se em dois estágios complementares: decomposição e reconstrução da série temporal. Cada estágio é composto por dois passos que formam os quatro passos da técnica (em inglês): *Embedding*, *Singular Value Decomposition (SVD)*, *Grouping* e *Diagonal Averaging*.

Decomposição

No estágio da decomposição, a série temporal inicial é decomposta em uma soma de poucas subséries, de modo que cada subsérie possa ser identificada e interpretada como componentes constitutivos.

- Primeiro passo: *Embedding*

Considere uma série temporal unidimensional real e não nula, i.e. com pelo menos um valor diferente de zero, $Y_t = Y_1, \dots, Y_N$, sendo N o comprimento da série ou a quantidade de observações ao longo do intervalo de tempo investigado.

Inicialmente, a série original unidimensional é transformada em uma série multidimensional com dimensão L , onde L é dito o comprimento da janela. Este é o único parâmetro deste passo e representa a quantidade de componentes em que a série é decomposta. Ele deve ser um valor inteiro, entre $2 \leq L \leq N$, e segundo resultados teóricos o tamanho de L deve ser suficientemente grande, mas não superior a $N/2$ (HASSANI, 2007).

A série temporal multidimensional, que é uma sequência de vetores constituídos por elementos da série Y_t , forma a matriz apresentada na expressão (1), denominada matriz trajetória, resultado desse primeiro passo (Golyandina, Nekrutkin e Zhigljavsky, 2001; Hassani, 2007; Hassani, Heravic e Zhigljavsky, 2009).

$$\begin{bmatrix} Y_1 & Y_2 & \cdots & Y_K \\ Y_2 & Y_3 & \cdots & Y_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ Y_L & Y_{L+1} & \cdots & Y_N \end{bmatrix} \quad (1),$$

em que, $K = N - L + 1$ é o número de vetores deslocados no tempo.

- Segundo passo: *Singular Value Decomposition (SVD)*

Neste passo, decomposição do valor singular, é realizada a decomposição da matriz trajetória X em uma soma de matrizes elementares.

Seja $S = X X^T$ e $\lambda_1 \geq \dots \geq \lambda_L \geq 0$ autovalores de S , com U_1, \dots, U_L , os correspondentes autovetores, formando um sistema ortonormal.

Representando os componentes principais da matriz trajetória como $V_i = \frac{X^T U_i}{\sqrt{\lambda_i}}$, $i = 1, 2, \dots, d$ a decomposição é representada como:

$$X = E_1 + \dots + E_d \quad (2),$$

onde d denota o número de autovalores diferentes de zero da matriz S e $E_i = \sqrt{\lambda_i} U_i V_i^T$ matrizes elementares.

O conjunto $(\sqrt{\lambda_i}, U_i, V_i)$ é conhecido como o i -ésimo *autotriple* da matriz trajetória \mathbf{X} e $\sqrt{\lambda_i}$ é o seu valor singular (Golyandina, Nekrutkin e Zhigljavsky, 2001; Hassani, 2007; Hassani, Heravic e Zhigljavsky, 2009).

Reconstrução

- Terceiro passo: *Grouping*

No terceiro passo ocorre a junção das matrizes elementares \mathbf{E}_i em vários grupos e a soma delas dentro de cada grupo. O passo “*grouping*” particiona o conjunto de índices da expressão (2) $(1, \dots, d)$ em subconjuntos disjuntos I_1, \dots, I_m , fornecendo a representação

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

Portanto, o resultado desse passo é a representação da matriz trajetória como uma soma de matrizes resultantes $(\mathbf{E}_{I_1}; \dots; \mathbf{E}_{I_m})$. A escolha dos conjuntos I_1, \dots, I_m é a segunda e última decisão necessária para a aplicação do método SSA. Essa escolha é baseada na propriedade denominada *separabilidade*. A separabilidade entre os conjuntos pode ser mensurada pela correlação ponderada, calculada da seguinte forma: a correlação ponderada entre duas subséries $Y_t^{(1)}$ e $Y_t^{(2)}$ pode ser expressa como,

$$\rho_{12}^{(w)} = \frac{\langle Y_t^{(1)}, Y_t^{(2)} \rangle_w}{\|Y_t^{(1)}\|_w \|Y_t^{(2)}\|_w};$$

em que a norma da i -ésima subsérie é dada por

$$\|Y_t^{(i)}\|_w = \sqrt{\langle Y_t^{(i)}, Y_t^{(i)} \rangle_w};$$

com o produto interno sendo definido por

$$\langle Y_t^{(i)}, Y_t^{(j)} \rangle = \sum_{c=1}^N w_c Y_c^{(i)} Y_c^{(j)}; i, j = 1, 2,$$

onde os pesos w_c são dados por $w_c = \min\{c, L, N-c\}$, e assume-se que $L \leq N/2$.

Em geral, avalia-se a correlação ponderada entre o *sinal*, que é o agrupamento dos principais autotriples, e o *ruido*, que é o grupo formado pelos autotriples remanescentes. Se o valor absoluto da correlação ponderada é muito pequeno, tem-se que as duas séries são quase ortogonais. Diz-se então que estes componentes são separáveis.

- Quarto passo: *Diagonal averaging*

A operação realizada neste último passo obtem para cada uma das matrizes resultantes uma aproximação para os componentes da série original, ou seja transforma cada matriz da decomposição agrupada (3) em uma nova série de tamanho N , que pode ser considerada como uma aproximação da série original.

2.2.2 Algoritmo recorrente de previsão

Seja $\tilde{Y} = \tilde{Y}_1 + \tilde{Y}_2 + \dots + \tilde{Y}_N$ a série reconstruída ou aproximada, conforme o segundo estágio da SSA. A previsão dos valores futuros \tilde{Y}_{N+h} , com $h = 1, \dots, M$, é obtida a partir da

seguinte expressão recorrente:

$$\hat{Y}_{N+h} = \sum_{p=1}^{L-1} a_p \hat{Y}_{(N+h)-p}; h = 1, \dots, M$$

Em que os a_p são os coeficientes da combinação linear entre os $L-1$ últimos termos da série reconstruída. É claro que, quanto maior o número de passos à frente (h), mais as previsões dependerão da qualidade das predições anteriores. Para que a predição considere pelo menos um valor aproximado da série original, é necessário um horizonte de no máximo $M = L-1$. Os números $\hat{Y}_{N+1}, \hat{Y}_{N+2}, \dots, \hat{Y}_{N+M}$ formam os M termos preditos pelo algoritmo recorrente de previsão, fundamentado no SSA básico.

Detalhes sobre o funcionamento do algoritmo estão disponíveis na obra da Golyandina, Nekrutkin e Zhigljavsky (2001).

3. Desenvolvimento dos estudos

3.1 Cenários

Visando comparar métodos propostos e recomendados na literatura de séries temporais com a metodologia SSA, foram definidos dois cenários, para os quais modelou-se e estimou-se valores futuros de uma série temporal real (estudo empírico) e uma série artificial (estudo simulado).

No estudo empírico considerou-se uma série meteorológica, com dados da velocidade média do vento. Esta série abrange o período entre 03 de setembro de 2010 às 12:00 e 12 de setembro de 2011 às 11:30, totalizando 17892 observações ao longo do período especificado, a cada meia hora. Considerou-se para o conjunto de teste as informações de 2688 instâncias no tempo. A coleta foi realizada em uma mini estação meteorológica, localizada na unidade CIMATEC do SENAI-BA, que se encontra instalada a uma altura de 36 metros em relação ao nível do solo. Esta série meteorológica será rotulada como *VENTO* neste trabalho.

Para o estudo simulado, uma série temporal foi gerada de acordo com o processo estocástico ARIMA ($p=1, d=1, q=2$). O parâmetro da parte regressiva foi fixado em 0,4 e os valores para os parâmetros de médias móveis foram $\theta_1 = 0,3$ e $\theta_2 = 0,8$. A escolha dos parâmetros fundamentou-se em exemplos do livro Morettin e Toloi (2006). Para esse processo foram geradas 212 observações, as doze últimas alocadas no conjunto de teste.

3.2 Medidas de precisão

A avaliação das capacidades preditivas dos métodos considerados foi conduzida em diferentes horizontes de previsão. Para a extensa série meteorológica, os seguintes horizontes à frente foram considerados: 1, 24, 48, 96, 192, 336, 672, 1008, 1344, 2016, 2688. Para as avaliações na série simulada foram considerados os 12 passos à frente. Os resultados das previsões foram comparados utilizando-se o *erro quadrático médio (MSE)* e o *erro percentual total (TPE)* e, é claro, o erro cometido na estimação para cada horizonte.

O erro quadrático médio (MSE) é definido pela expressão abaixo:

$$MSE = \frac{1}{n} \sum_{j=k+1}^T e_j^2$$

em que $e_j = Y_j - \hat{Y}_j$ são os erros de previsão (\hat{Y}_j é a previsão de Y_j) e n é a quantidade de

observações reservadas para o conjunto de teste, com $k = T - n$.

O erro percentual total (TPE), fornece uma medida relativa do erro. O TPE é calculado como

$$TPE = \left(\frac{\sum_{j=k+1}^T e_j}{\sum_{j=k+1}^T Y_j} \right) \times 100.$$

3.3 Processo de modelagem

3.3.1 Técnicas clássicas

Conforme veremos a seguir, a série empírica (VENTO) não é estacionária, exibindo tendência e sazonalidade. Usaremos o algoritmo de Holt-Winters, na forma aditiva porque a série possui valores nulos. Para o método de Box-Jenkins, o mais adequado para modelar as variações na série meteorológica foi o SARIMA. A série empírica precisou de apenas uma diferença para se tornar um processo integrado [I(1)] e assim especificou-se a classe do modelo SARIMA (p,1,q)×(P,D,Q). A seleção do modelo mais adequado para representar o processo gerador da série baseou-se no critério AIC (critério de informação de Akaike; Akaike, 1973, 1974).

Na modelagem clássica para a série artificial foi utilizado o alisamento exponencial de Holt (SEH) e um modelo correspondente, da classe ARIMA. Utilizou-se a SEH ao invés do Holt-Winters, porque a série simulada não apresenta comportamentos sazonais. O modelo ARIMA foi considerado como um “padrão ouro”, isto é, o ARIMA serviu como medida para comparação dos outros métodos adotados, na medida em que se espera que a predição realizada nos dados oriundos do processo estocástico puro, tenha maior precisão do que a feita com as demais técnicas.

3.3.2 SSA

Como foi discutido na Seção 2.2.1, a modelagem via SSA é feita mediante a escolha dos dois únicos parâmetros, a saber, o comprimento da janela L (no estágio de decomposição) e o processo estrutural de agrupamento, na reconstrução da série temporal.

Na análise via SSA considerou-se o comprimento ótimo da janela ($L_{\text{ótimo}}=N/2$) para a série artificial e para a série VENTO, por dificuldades computacionais, baseamos a escolha do valor de L no período sazonal da série, ou seja escolheu-se um comprimento múltiplo da sazonalidade.

A separação entre o *signal* e o *ruído* na fase de reconstrução das séries foi avaliada de acordo com os valores singulares e a correlação ponderada. Além dos critérios ora citados, foi usado também o percentual acumulado de explicabilidade dos possíveis componentes formadores do *signal*.

4. Resultados

Estudo empírico

As características da série temporal VENTO podem não ser identificadas facilmente através do seu gráfico temporal. Assim como apresentado na Figura 1, nota-se uma grande massa de dados ao longo dos instantes de tempo.

Para avaliar a suposição de estacionariedade foi usado o bem conhecido teste KPSS (Kwiatkowski *et. al.*, 1992), que rejeitou a hipótese (p -valor $< 0,01$). Além disso, os testes de Box-Pierce (Box e Pierce, 1970) e de Ljung-Box (Ljung e Box, 1978) apontaram a existência de

autocorrelação conjunta, estatisticamente significativa ($p < 2,2 \times 10^{-16}$). De acordo com esses testes temos indícios de que a série *VENTO* não é estacionária.

Com o intuito de facilitar a visualização da periodicidade da série, considerou-se um recorte da série temporal *VENTO*. A Figura 1(b) apresenta uma ampliação do componente sazonal, com o comportamento do componente periódico para os 336 instantes de tempo iniciais. Notamos uma repetição sazonal na série, repetição com uma amplitude igual a 48, que corresponde aos 48 instantes de medição efetuadas por dia.

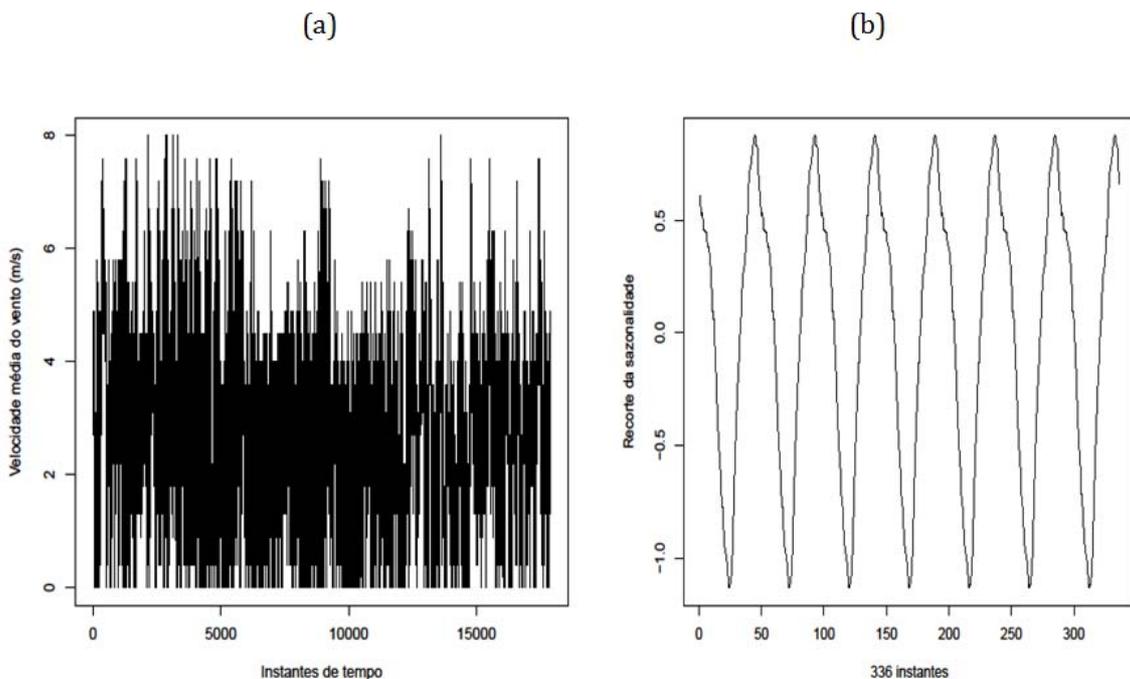


Figura 1: Velocidade média do vento ao longo do tempo (a). Comportamento sazonal para os primeiros 336 instantes de tempo (b).

Modelagem

Os dois modelos de Box-Jenkins que apresentaram menores valores no critério AIC foram SARIMA (2,1,2)x(1,1,1), que rotulamos como SARIMA 1 (AIC igual a 35394,05) e SARIMA (3,1,3)x(1,1,1), identificado como SARIMA 2 (AIC igual a 35389,52).

Para a modelagem via SSA, usamos um comprimento de janela $L = 48 \times 84 = 4032$. Ao analisar as correlações ponderadas nota-se que a partir do componente de posto 200 seguem-se muitos componentes com correlações baixas, os quais irão compor o grupo do *ruído*. Pela análise do comportamento dos principais autovetores, notou-se um comportamento periódico representado pela formação dos pares de autovetores de postos (2, 3) e (4, 5).

Ao finalizar a análise dos estágios da SSA (decomposição e reconstrução), aplicou-se o algoritmo recorrente de previsão SSA na série reconstruída a partir dos 1000 primeiros componentes da SVD, por motivos computacionais.

Previsão

A Tabela 1 apresenta um comparativo geral dos resultados de previsão de cada método analisado. Essa tabela resume a acurácia das técnicas, apresentando o erro quadrático médio (MSE) e o erro percentual total (TPE).

Tabela1: Comparação entre as técnicas consideradas, para a série (VENTO).

Horizonte	HW aditivo		SARIMA 1		SARIMA 2		SSA	
	MSE	TPE (%)	MSE	TPE (%)	MSE	TPE (%)	MSE	TPE (%)
1	0,170	-13,285	0,137	-11,948	0,144	-12,250	0,927	-31,065
24	1,242	18,404	1,629	24,884	1,626	25,088	1,117	6,597
48	0,944	6,917	1,312	12,250	1,323	11,763	1,301	-3,244
96	0,819	13,203	1,150	18,211	1,112	17,269	1,208	3,976
192	2,693	28,372	3,200	32,619	3,064	31,584	3,296	26,668
336	4,608	36,790	5,361	40,719	5,146	39,669	5,201	35,471
672	3,529	21,609	3,874	27,027	3,760	25,500	4,053	23,109
1008	2,889	15,713	3,103	22,126	3,012	20,293	3,262	19,798
1344	2,441	12,424	2,603	19,690	2,522	17,603	3,100	17,988
2016	2,891	14,081	3,222	22,373	3,091	19,984	3,721	22,009
2688	2,924	13,842	3,333	23,295	3,168	20,573	3,881	22,748

Observa-se que, de acordo com o MSE, o método de Holt-Winters sazonal aditivo foi superior aos demais, pois para quase todos os horizontes, os MSE's para este método foram inferiores ao SARIMA 1, SARIMA 2 e SSA. Excetua-se o passo $h=24$, para o qual a SSA se mostrou superior. Se a avaliação utiliza o erro relativo, TPE, o algoritmo Holt-Winters e SSA vencem em igual quantidade de horizontes. O algoritmo Holt-Winters apresentou melhor desempenho em horizontes maiores ($h=672, 1008, 1344, 2016, 2688$) enquanto que para previsões de curto e médio prazo ($h=24, 48, 96, 192$ e 336) a previsão SSA apresentou uma melhor performance.

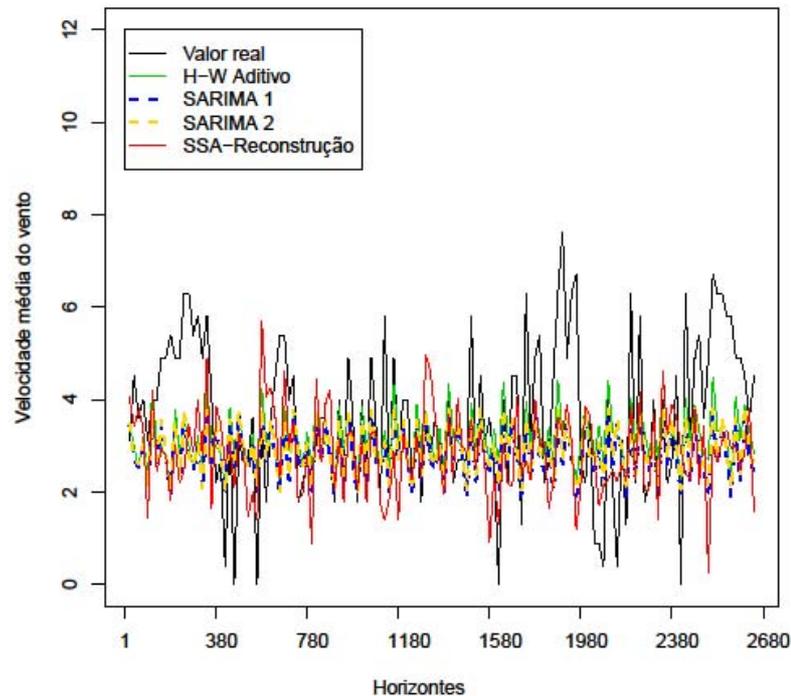


Figura 2: Previsões dos métodos comparadas com os valores reais (VENTO).

Nota-se na Figura 2 que o algoritmo recorrente de previsão SSA conseguiu detectar melhor as flutuações das velocidades médias do vento [“valor real”] em comparação ao Holt-Winters aditivo e os modelos ARIMA sazonais retratando melhor o conjunto de teste. Nota-se também que o algoritmo Holt-Winters resultou em previsões melhores do que os modelos ARIMA sazonais, uma vez que o segmento correspondente (em verde) possui uma amplitude um pouco maior do que as observadas nos segmentos em azul (SARIMA 1 e SARIMA 2). A ordem de performance dos métodos escolhidos foi portanto SSA (com componentes 1 a 1000) seguida do Holt-Winters aditivo, depois SARIMA (3,1,3)x(1,1,1) e por fim SARIMA (2,1,2)x(1,1,1).

Estudo simulado

O gráfico com a estrutura do processo ARIMA simulado, Figura 3(a) e com o logaritmo dos autovalores resultantes da decomposição SSA, Figura 3(b), estão apresentados abaixo.

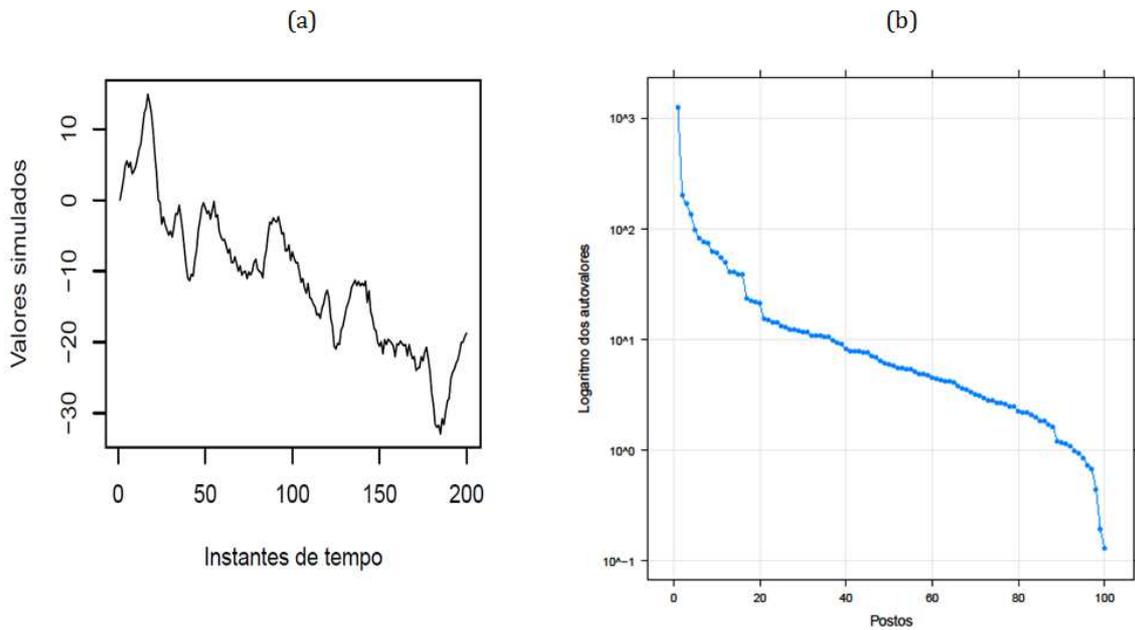


Figura 3: Processo ARIMA (a).

Logaritmo dos autovalores (b).

Através do gráfico (a) da Figura 3, observa-se que o processo ARIMA simulado indica uma clara tendência decrescente com algumas flutuações, que aparentemente não podem ser qualificadas como sazonais. O gráfico (b) mostra o espectro da matriz trajetória utilizada na decomposição SSA. Podemos notar que a decomposição desta série proveniente de um processo ARIMA, indica claramente um componente de tendência, representado pelo primeiro autovalor em destaque. Nota-se também que o gráfico (b) exibe três grandes fases ao longo dos 100 postos dos autovalores. Ao investigarmos a primeira fase, se considerarmos um agrupamento consistindo de somente os 16 primeiros autovalores como *signal* e os demais componentes como *ruído*, observamos uma correlação ponderada entre o *signal* e o *ruído* de 0,0069, sendo que o *signal* explica 99,66% da variação da série. Se tomamos os 20 primeiros autovalores, observamos uma correlação ponderada entre o *signal* e o *ruído* de 0,0039, sendo que o *signal* explica agora 99,77% da variação da série. Ao usarmos os 88 primeiros autovalores obtemos como resultados 0,0017 e 99,99% respectivamente.

De posse dessas opções de agrupamento aplicou-se o algoritmo recorrente de previsão na série simulada, confrontando as previsões geradas a partir dos diferentes espaços trajetória. Os melhores resultados foram obtidos utilizando-se os 88 primeiros autovalores, um resultado já antecipado, ou seja, para a série originária do processo ARIMA (1,1,2) o melhor resultado preditivo observado corresponde ao *signal* formado pelos 88 primeiros autotriples.

A Tabela 2 apresenta os valores futuros (rotulados na tabela como “Real”), as diferenças entre os valores originais do conjunto de teste e os previstos por cada técnica (nomeados como “Erro”), além das medidas MSE e TPE. Pode-se observar que houve superestimação das predições em todos os horizontes e métodos avaliados.

Ao analisar o MSE nota-se que a liderança do “padrão ouro” permaneceu nos horizontes $h = 2, 3, 4, 5$ ficando em segundo lugar nos passos h de 6 a 12. Para esses últimos horizontes a SSA se revelou melhor e ficou em segunda posição no passo 5.

Ao avaliar o erro relativo, o desempenho agora foi semelhante ao observado para o MSE. Nota-se que o “padrão ouro” apresenta a melhor performance em 3 dos horizontes (2, 3 e 4) e fica com a segunda posição para os passos restantes. Por outro lado, a SSA mostrou-se melhor em 8 dos horizontes ($h =$ de 5 a 12) e foi o segundo melhor resultado nos passos 2 e 4.

Tabela 2: Comparação entre as previsões (Processo teórico ARIMA).

h	Real	SEH			ARIMA(1,1,2)			SSA		
		Erro	MSE	TPE	Erro	MSE	TPE	Erro	MSE	TPE
1	-18,579	-18,579	0,103	1,729	-18,579	0,318	-3,035	-18,579	3,196	9,623
2	-18,821	-18,821	0,568	3,576	-18,821	0,255	-0,337	-18,821	2,838	0,569
3	-18,377	-18,377	0,729	4,237	-18,377	0,199	0,301	-18,377	4,069	4,963
4	-20,244	-20,244	3,344	7,509	-20,244	1,447	3,218	-20,244	10,510	-3,543
5	-22,230	-22,230	9,366	11,697	-22,230	4,874	6,878	-22,230	8,513	-3,478
6	-24,278	-24,278	19,245	16,141	-24,278	10,840	10,720	-24,278	9,734	-6,037
7	-26,532	-26,532	33,757	20,642	-26,532	19,953	14,607	-26,532	8,976	-6,375
8	-28,318	-28,318	51,417	24,805	-28,318	31,049	18,153	-28,318	11,708	-8,487
9	-29,625	-29,625	70,671	28,497	-29,625	42,901	21,225	-29,625	13,320	-4,799
10	-30,217	-30,217	89,316	31,626	-30,217	53,807	23,718	-30,217	12,035	-4,477
11	-28,670	-28,670	101,490	33,835	-28,670	59,482	25,215	-28,670	16,880	-0,955
12	-29,909	-29,909	116,087	36,037	-29,909	66,565	26,729	-29,909	21,600	2,040

O gráfico da Figura 4 mostra uma comparação entre os métodos preditivos selecionados e o conjunto de teste, para o processo estocástico ARIMA.

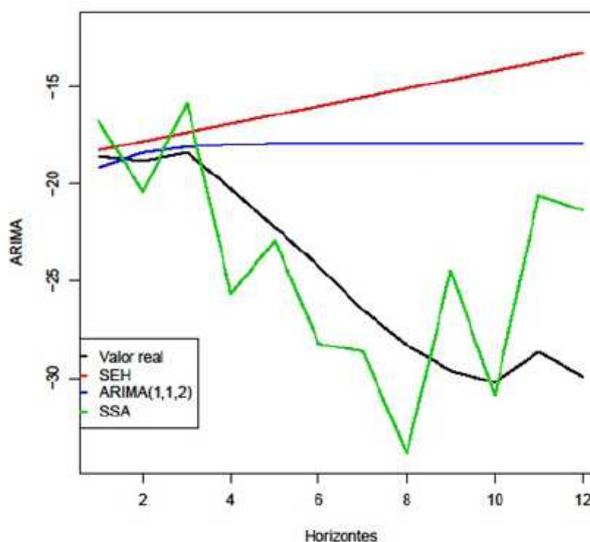


Figura 4: Comparação das previsões (Processo ARIMA).

Observa-se graficamente que o “padrão ouro” e o alisamento exponencial de Holt (SEH) geraram previsões quase em linha reta, simplificando em demasia o comportamento real dos valores futuros, sendo que para o “padrão ouro” as previsões foram quase constantes. Os resultados obtidos pela SSA por outro lado acompanham a curva do conjunto de teste embora exibam mais variações do que as apresentadas pelos dados reais.

5. Considerações finais

O presente artigo teve como objetivo apresentar a metodologia SSA e fazer uma breve avaliação da sua capacidade preditiva em duas séries, uma empírica e outra simulada,

confrontando-a com alguns importantes métodos clássicos, a saber, a suavização exponencial de Holt (SEH), a suavização exponencial de Holt-Winters (H-W) e a classe de modelos de Box e Jenkins. No contexto do estudo empírico, verificou-se que a SSA mostrou-se melhor nas previsões à curto e médio prazo, levando em consideração o erro relativo (TPE). Por outro lado, utilizando-se o MSE, o algoritmo de Holt-Winters apresentou melhor desempenho preditivo.

No estudo da série simulada, observou-se que tanto pelo MSE quanto pelo TPE, a SSA exibiu previsões mais precisas para o processo ARIMA (1,1,2), um processo bastante comum em séries reais.

De uma forma geral, nota-se que o algoritmo recorrente de previsão SSA consegue representar melhor as variações existentes nos dados, como flutuações sazonais e picos, características encontradas com frequência em séries históricas. A previsão SSA apresentou assim um comportamento global mais condizente com a realidade das séries.

Uma grande vantagem em aplicar a SSA em substituição aos modelos de Box-Jenkins refere-se à sua simplificação no entendimento das séries temporais e conseqüente diminuição da intervenção do analista, pois o processo de modelagem e previsão via SSA leva em consideração uma fundamentação teórica que envolve a decomposição e reconstrução dos componentes constitutivos da série temporal, sem prejuízo da sua capacidade preditiva. Sua utilização produziu resultados tão bons ou mesmo superiores aos gerados pelos métodos clássicos considerados neste artigo.

Referências

- Akaike, H.** *Maximum likelihood identification of Gaussian autoregressive moving average models.* Biometrika, 60, 255-265, 1973
- _____. *A new look at the statistical model identification.* IEEE Transactions on Automatic Control, AC-19, 716-723, 1974
- Box, G.E.P. e Jenkins, G.M.** *Time Series Analysis: Forecasting and Control.* São Francisco: Holden-Day, 1970.
- Box, G.E.P. and Pierce, D.A.** *Distribution of residual correlations in autoregressive-integrated moving average time series models.* Journal of the American Statistical Association, 65, 1509-1526, 1970
- Brillinger, D.R.** *Time Series: Data analysis and Theory.* Classics in Applied Mathematics. SIAM, 2001.
- Brockwell, P.J. e Davis, R.A.** *Introduction to Time Series and Forecasting.* 2nd ed. New York: Springer Texts in Statistics, 2002.
- Golyandina, N., Nekrutkin, V. e Zhigljavsky, A.** *Analysis of time series structure: SSA and related techniques.* New York: Chapman & Hall/CRC, 2001.
- Hassani, H.** *Singular Spectrum Analysis: Methodology and Comparison.* Journal of Data Science 5, 239-257, 2007.
- Hassani, H., Heravic, S. e Zhigljavsky, A.** *Forecasting European industrial production with SSA.* International Journal of Forecasting 25, 103-118, 2009.
- Kwiatkowski, D.; Phillips, P.C.B.; Schmidt, P. and Shin, Y.** *Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root.* Journal of Econometrics, 54, 159-178, 1992.
- Ljung, G.M. and Box, G.E.P.** *On a measure of lack of fit in time series models.* Biometrika, 65, 297-303, 1978.
- Morettin, P.A. e Toloi, C.M.C.** *Análise de séries temporais.* São Paulo: Edgard Blucher/ ABE-Projeto Fisher, 2006.