

Victor Rodrigues Corrêa

**Abordagem Singular Spectrum Analysis na  
Modelagem de Séries Temporais de  
Velocidade do Vento de Curtíssimo Prazo**

Niterói - RJ, Brasil

07 de julho de 2015

**Victor Rodrigues Corrêa**

**Abordagem Singular Spectrum  
Analysis na Modelagem de Séries  
Temporais de Velocidade do Vento  
de Curtíssimo Prazo**

**Trabalho de Conclusão de Curso**

Monografia apresentada para obtenção do grau de Bacharel em  
Estatística pela Universidade Federal Fluminense.

Orientador: Prof. Moisés Lima de Menezes

Niterói - RJ, Brasil

07 de julho de 2015

**Victor Rodrigues Corrêa**

**Abordagem Singular Spectrum Analysis na  
Modelagem de Séries Temporais de  
Velocidade do Vento de Curtíssimo Prazo**

Monografia de Projeto Final de Graduação sob o título “*Abordagem Singular Spectrum Analysis na Modelagem de Séries Temporais de Velocidade do Vento de Curtíssimo Prazo*”, defendida por Victor Rodrigues Corrêa e aprovada em 07 de julho de 2015, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

---

**Prof. Dr. Moisés Lima de Menezes**  
Departamento de Estatística – UFF

---

**Profa. Dra. Keila Mara Cassiano**  
Departamento de Estatística – UFF

---

**Profa. Dra. Márcia Marques de Carvalho**  
Departamento de Estatística – UFF

M 149 Corrêa, Victor Rodrigues  
Abordagem Singular Spectrum Analysis na  
Modelagem de Séries Temporais de Velocidade  
do Vento de Curtíssimo Prazo/Victor Rodrigues  
Corrêa. - Niterói: [s. n.], 2015.  
37f.

Trabalho de Conclusão de Curso - (Bacharelado em Es-  
tatística) – Universidade Federal Fluminense, 2015.

1. Singular Spectrum Analysis. 2. Modelagem.  
3. Velocidade do Vento.  
I. Título.

CDD. 519.536

# Resumo

A questão energética está se tornando cada vez mais importante no mundo. Com o crescimento da qualidade de vida e da industrialização, a energia se torna cada vez mais indispensável para o dia a dia da população e para o crescimento econômico. Diante disto, surge uma questão: Como suprir essas necessidades energéticas, sem agredir os recursos naturais? Uma alternativa viável é a utilização da energia hídrica como fonte. Porém, com o aumento populacional e as frequentes mudanças climáticas que proporciona grandes períodos de estiagem, percebe-se que outra forma de geração de energia renovável seria fundamental para complementar a energia hidrelétrica nestas épocas. Assim a energia eólica entrou em ação, pois além de ser uma energia limpa, renovável e não utilizar, de maneira predatória, os recursos naturais, é uma energia criada a partir do vento, ou seja, em tempos de estiagem, pode haver uma tendência de aumento das correntes do vento, proporcionando esta complementariedade. Desta forma, a modelagem e previsão de séries temporais de velocidade do vento é essencial para o planejamento de uso deste tipo de geração de energia como complemento à hídrica. Séries temporais filtradas podem gerar modelos mais adequados para previsão. Este projeto tem por finalidade modelar uma série temporal de velocidade do vento de curtíssimo prazo a partir de duas conhecidas classes de modelos: Os modelos de Holt-Winters e de Box & Jenkins. Além disso, pretende-se realizar uma filtragem na série original através da abordagem Singular Spectrum Analysis (SSA), removendo o seu ruído, e modelar também a série filtrada menos ruidosa via Holt-Winters e Box & Jenkins. Os modelos, tanto com a série original como com a série filtrada, são comparados através de estatísticas de aderências, de modo que o melhor modelo será aquele que minimize os erros. Os resultados mostram que os modelos de Box & Jenkins apresentam uma maior capacidade preditiva e que a filtragem melhora muito esta capacidade.

Palavras-chaves: Singular Spectrum Analysis, Modelagem, Velocidade do Vento, Filtragem, Séries Temporais.

# Agradecimentos

Sou grato primeiramente ao meu Deus, por ter sido minha rocha e a minha força nos tempos difíceis desta graduação. Todo sustento e capacidade vieram dEle. Agradeço também pelo apoio recebido pela minha tia Maria Isaura, que nunca mediu esforços para realizar meus sonhos e superar todas as adversidades, em toda minha vida acadêmica. Ela sempre será a minha inspiração. Faço ainda um agradecimento especial a minha melhor amiga: A minha noiva. Seu apoio e companhia tornaram os anos mais difíceis da minha vida nos mais felizes também. Agradeço também aos meus amigos, que sempre estiveram ao meu lado me apoiando mesmo quando a graduação me fazia ficar ausente em alguns momentos. Quero agradecer muito ao professor orientador deste trabalho, Prof. Moisés Lima De Menezes, que não só me orientou de maneira incrível, como acreditou em mim, e soube resgatar a minha confiança. Por fim, quero agradecer a Leonora e todos os conselheiros do CEP, onde fui muito bem acolhido, quando precisei, e me deram todo respaldo para o termino da minha graduação. .

# Sumário

## Lista de Figuras

## Lista de Tabelas

<b>1</b>	<b>Introdução</b>	p. 10
1.1	Contextualização . . . . .	p. 10
1.2	A Energia Eólica no Brasil e no Mundo . . . . .	p. 11
1.3	Modelos de Previsão para Velocidade do Vento . . . . .	p. 12
1.4	Filtragem de Séries Temporais . . . . .	p. 13
1.5	Estrutura . . . . .	p. 13
<b>2</b>	<b>Objetivos</b>	p. 14
<b>3</b>	<b>Materiais e Métodos</b>	p. 15
3.1	Banco de Dados . . . . .	p. 15
3.2	Componentes de uma Série Temporal . . . . .	p. 15
3.3	Modelagem de Séries Temporais . . . . .	p. 17
3.3.1	Modelos de Holt-Winters . . . . .	p. 17
3.3.2	Modelos de Box & Jenkins . . . . .	p. 19
3.4	Filtragem SSA . . . . .	p. 23
3.4.1	Decomposição . . . . .	p. 23
3.4.2	Reconstrução . . . . .	p. 24
3.5	Resumo da Metodologia . . . . .	p. 26

<b>4</b>	<b>Análise dos Resultados</b>	p. 28
4.1	Filtragem SSA . . . . .	p. 28
4.2	Modelagem . . . . .	p. 31
<b>5</b>	<b>Conclusão</b>	p. 34
	<b>Referências</b>	p. 36



# Lista de Figuras

1	Fluxograma da metodologia. . . . .	p. 26
2	Série original de velocidade do vento medida a cada 10 minutos. . . . .	p. 28
3	Os 9 primeiros vetores singulares e suas respectivas contribuições à série. . . . .	p. 29
4	Gráfico de dispersão de 6 pares dos autovetores. . . . .	p. 29
5	As três Componentes obtidas no Agrupamento SSA. . . . .	p. 30
6	Comparação gráfica entre a Série Original e a Série Filtrada SSA. . . . .	p. 31

# Lista de Tabelas

1	Equações Básicas dos modelos de Holt-Winters . . . . .	p. 18
2	Características principais dos modelos $AR(p)$ , $MA(q)$ e $ARMA(p, q)$ . . . . .	p. 21
3	Agrupamento dos Vetores Singulares após a Análise Gráfica. . . . .	p. 30
4	Correlação Ponderada entre as três Componentes SSA. . . . .	p. 31
5	Estatísticas de Aderência aplicada aos Modelos Testados. . . . .	p. 32
6	Estatísticas de Aderência aplicada aos Modelos Testados (In Sample). . . . .	p. 33
7	Estatísticas de Aderência aplicada aos Modelos Testados (Out of Sample). . . . .	p. 33

# 1 Introdução

## 1.1 Contextualização

A questão energética tem sido um dos assuntos de maior importância no mundo. O aumento da qualidade de vida da sociedade está diretamente ligado ao consumo de energia. Assim como o aumento na demanda energética em função da melhoria dos padrões de vida nos países em desenvolvimento. Esta questão gera preocupações no planejamento energético de todas as economias emergentes. Dentre elas, podemos citar a segurança no suprimento de energias necessária para o desenvolvimento social e econômico de um país e os custos ambientais para esse consumo de energia [7].

Para suprir esse crescimento mundial do consumo de energia, é necessária uma grande dependência das sociedades industriais modernas dos combustíveis fósseis, principalmente o petróleo. Os derivados do petróleo, assim como o carvão, são os maiores responsáveis pela produção de gases causadores do efeito estufa, além de não possuírem reservas permanentes. A dependência destes ameaça a estabilidade ecológica do planeta e deixa a sociedade em uma situação desconfortável em relação a oferta de energia para o futuro. Outra forma de geração de energia é a nuclear, esse tipo de geração de energia elimina o problema de emissões de gases na atmosfera. Porém, este benefício é substituído pelos resíduos nucleares, que sem solução até hoje encontrada, representam uma terrível herança deixada através dos séculos às futuras gerações [3].

Para que essa demanda de energia seja atendida, sem que o planeta sofra tanto ecologicamente, são necessárias outras formas menos agressivas para a geração de energia, menos poluentes, conservando os recursos naturais [3].

Existem alternativas para reduzir a dependência das fontes convencionais, como as fontes renováveis de geração de energia tais como: hídrica, eólica, solar, biomassa, etc. Dentre elas destaca-se a fonte hídrica, utilizada em várias regiões do mundo [3].

Segundo[5], mais de 80% da energia gerada no Brasil é de origem hidrelétrica, Mesmo

sendo uma tecnologia considerada limpa, essa maneira de produção de energia tem a necessidade de represar os rios, causando um desequilíbrio ambiental e conseqüentemente gerando impacto no meio ambiente e na população ribeirinha.

Nos períodos de estiagem, a geração hidrelétrica de energia pode ser prejudicada, tornando imprescindível a necessidade de uso de outra fonte de energia para complementar a energia hidrelétrica. De acordo com [6], os ventos possuem maior intensidade na primavera. Este período do ano coincide com o período em que os reservatórios de água estão em baixo nível. Portanto, a energia eólica poderia ser uma geração alternativa de energia para suprir a necessidade em nestes períodos.

Pela sua grande extensão o Brasil apresenta diferentes climas da região norte até a Região Sul [1]. Entre os grandes fatores que influem no clima brasileiro estão a Zona de Convergência Intertropical ao norte para onde convergem os ventos alísios; as distintas ações exercidas pelo relevo continental; e a ação periódica irregular das massas de ar polares que adentram as regiões sul e sudeste em maior intensidade. O clima tropical que abrange toda a costa que vai desde o Maranhão até partes de São Paulo, estendendo-se a oeste até o Mato Grosso e Mato Grosso do Sul passando pelo nordeste e chegando até os estados centrais como Goiás e Tocantins. Este clima é caracterizado por chuvas com sazonalidades bem definida: estação seca de 4-5 meses (abril-setembro) e chuvosa (novembro-março). Essa sazonalidade tem importante papel energético no sistema elétrico brasileiro, de geração predominantemente hidráulica [1].

## 1.2 A Energia Eólica no Brasil e no Mundo

Para entender energia eólica, é necessário saber primeiramente o seu agente propulsor que é o vento. Os ventos têm origem na dinâmica da atmosfera terrestre e são causados, principalmente, pelos gradientes de pressão da atmosfera aliados ao movimento próprio de rotação da terra. Os gradientes de pressão são causados pelos gradientes da temperatura resultantes do aquecimento diferenciado da superfície da terra pelo sol. A região próxima da linha do equador, com latitudes próximas de zero, recebem uma maior incidência de energia proveniente do sol em relação aos trópicos e aos polos da terra, dando origem aos gradientes da temperatura causando o movimento da atmosfera da terra. Então, os ventos e a energia eólica são efeitos permanentes da dinâmica do nosso planeta e por isso este tipo de geração é considerada uma fonte de energia renovável e limpa [3].

Segundo [6], o vento tem comportamentos variados de velocidade ao longo do tempo,

esse fator somado as características topográficas de uma região influenciam diretamente na velocidade do vento. Por isso, em certas regiões pode haver diferentes velocidades do vento. Desta forma, deve-se estudar a região antes da instalação de um aerogerador pois deve ser feito para que sejam levados em conta todos os fatores regionais que influenciam nas condições do vento.

Devido a sua variabilidade, a velocidade do vento é uma das características de maior importância na geração eólica. Em boa parte dos casos pode haver uma variação de até 10% entre o valor médio anual e a média ao longo dos anos, o que acontece também ao longo do ano, porém essa variação é sazonal, apresentando o mesmo comportamento ao longo do ano em períodos iguais, mas em anos diferentes. O vento também manifesta variações importantes ao longo do dia [6].

Por volta de 1930, nos EUA, iniciou-se a utilização da energia eólica para gerar energia elétrica com pequenos aerogeradores ligados a bateria, que possibilitaram o acesso à energia para habitantes afastados de centros urbanos, porém a energia eólica foi perdendo espaço com o passar dos anos em virtude da chegada da eletricidade via cabos até as áreas mais afastadas. A energia eólica ganhou um impulso nas décadas de 1970 e 1980 após a primeira crise do petróleo. Vários países europeus e os Estados Unidos, passaram a despende recursos em estudos de novas formas de geração de energia, onde a energia eólica foi incluída. Na década de 1980, os parques eólicos de grande escala, ligados diretamente a rede elétrica, já completavam a produção de energia. O Brasil também entrou nessa busca, com um projeto conjunto entre os institutos de pesquisas aeroespacial de Brasil e da Alemanha o que permitiu a criação da turbina de 100 kw [1].

### 1.3 Modelos de Previsão para Velocidade do Vento

O estudo de previsão de velocidade do vento tem importância fundamental para geração eólica de energia elétrica. Segundo [15], a dificuldade para a previsão são as variações da velocidade dos ventos. Dependendo da aplicação os modelos de previsão mais utilizados são:

- Modelo Físico;
- Modelos Estatísticos (análise de séries temporais);
- Modelo Combinado (Estatístico e Físico).

O foco deste trabalho está na previsão utilizando o modelo estatístico.

Dentre as classes de modelos estatísticos, estão os modelos de Box & Jenkins e os

modelos de Holt-Winters [12].

## 1.4 Filtragem de Séries Temporais

De acordo com [15], a realização de filtrações em séries temporais pode aumentar o poder preditivo do modelo. Neste trabalho será usado o método “Singular Spectrum Analysis” (SSA) como técnica de filtração. SSA é um poderoso método para análise e previsão de séries temporais e incorpora elementos de análise de séries temporais clássica, estatística multivariada, geometria multivariada, sistemas dinâmicos e processamento de sinais [8]. SSA pode ser aplicados em vários campos da matemática e da física à economia e matemática financeira, meteorologia e oceanografia para ciências sociais [10].

## 1.5 Estrutura

Este trabalho está organizado em cinco capítulos, sendo o primeiro a introdução, mostrando a evolução da energia no Brasil e no mundo, a importância de se ter uma energia limpa e sustentável para complementar a energia hidrelétrica em períodos de estiagem e a importância de fazer uma previsão da velocidade do vento. No capítulo 2 são apresentados os objetivos. Na sequência, no capítulo 3 são apresentados materiais e métodos utilizados neste trabalho. No capítulo 4 estão as análises dos resultados com discussões e no capítulo 5 estão resumidas as principais conclusões do trabalho bem como as sugestões e análises para futuros estudos sobre o tema.

## 2 Objetivos

Este trabalho tem como proposta fazer previsões da velocidade do vento de curtíssimo prazo, propondo uma filtragem SSA (Singular Spectrum Analysis) antes de modelar estatisticamente a série.

Após uma análise descritiva do banco, será modelado uma série de velocidade do vento com periodicidade de 10 minutos durante uma semana via Holt-Winters e Box-Jenkins com e sem a filtragem SSA. Ao final das modelagens, testar qual modelo tem melhor desempenho em relação à capacidade preditiva através de medidas estatísticas como MAPE (Mean Absolute Percentage Error), RMSE (Root Mean Squared Error) e os critérios AIC (Akaike Information Criterion) e BIC (Bayesian Information Criterion). Com isso, mostrar que a filtragem SSA melhora a capacidade preditiva do modelo e apresentar o melhor modelo para estudos futuros sobre a viabilidade de implementação de forma mais incisiva a energia eólica como fonte de energia complementar a energia hidrelétrica.

## 3 Materiais e Métodos

### 3.1 Banco de Dados

Para o estudo, considerou-se a série temporal da velocidade do vento, de curtíssimo prazo, medida a cada 10 minutos a uma altura de 50 metros. Os dados considerados compreendem o período de 01 a 07 de agosto do ano de 2009, da cidade de Florianópolis - SC. Os dados foram disponibilizados pela Rede SONDA (Sistema de Organização Nacional de Dados Ambientais) do INPE (Instituto Nacional de Pesquisas Espaciais). As informações estão disponíveis na página oficial do órgão: <http://sonda.ccst.inpe.br/>. A amostra considerada na estimação do modelo contém 1008 observações.

### 3.2 Componentes de uma Série Temporal

Define-se série temporal como sendo um conjunto de dados de observações ordenadas dado um parâmetro de tempo e com dependência de série, sendo esse espaço de tempo: horário, minutos, diário, semanal, mensal, trimestral, anual, etc [12].

Para que uma série seja classificada como uma série temporal é necessário que os dados apresentem uma dependência serial entre eles. Por exemplo: Os dados de uma variável aleatória  $Z$  (sendo  $Z =$  Consumo de energia) no instante  $t$  (sendo  $t$  o tempo), com  $1 \leq t \leq T$ , possa conter informações necessárias para que seja determinado o valor dessa variável no instante  $t + 1$ . Cabe dizer que  $T =$  Número de observações da série temporal estudada [15].

Um das suposições mais frequentes que se faz a respeito de uma série temporal é que ela é estacionária, ou seja, ela se desenvolve aleatoriamente no decorrer do tempo ao redor de uma média constante. Uma série pode ser não-estacionária, mas em um determinado período curto pode ser estacionária [12].

De acordo com [12], pode-se considerar as observações  $Z_t, t = 1, \dots, T$  de uma série



temporal. Um modelo de decomposição consiste em escrever  $Z_t$  como a soma de três componentes não-observáveis. A estrutura da série temporal é aditiva se  $S_t$  independe de  $T_t$  e multiplicativa se existir uma dependência entre essas duas componentes de acordo com (3.1) e (3.2).

$$Z_t = T_t + S_t + a_t, \quad (3.1)$$

$$Z_t = T_t \cdot S_t \cdot a_t, \quad (3.2)$$

onde,  $T_t$  é a tendência da série, que são aquelas que produzem mudanças graduais ao longo do prazo,  $S_t$  é a componente sazonal e  $a_t$  é um componente ruidosa. Os componentes  $T_t$  são normalmente provocadas, por exemplo, pelo crescimento da população, produto interno bruto, ou por outros fatores que falham na tentativa de produzir mudanças repentinas, mas produzem variações graduais e regulares ao longo do tempo [15].

Considerando uma série sem a componente sazonal  $Z_t$  o modelo considerado será como em (3.3) e (3.4):

$$Z_t = T_t + a_t, \quad (3.3)$$

$$Z_t = T_t \cdot a_t. \quad (3.4)$$

Há vários métodos para estimar  $T_t$ . Os mais utilizados consistem em:

- Ajustar a função do tempo, como um polinômio, uma exponencial ou outra função suave;
- Suavizar (ou filtrar) os valores da série ao redor de um ponto, para estimar a tendência naquele ponto;
- Suavizar os valores da série através de sucessivos ajustes de retas de mínimos quadrados ponderados.

Estimando a tendência através de  $T_t$  podemos obter a série ajustada para tendência

ou livre de tendência.

$$Y_t = Z_t - T_t. \quad (3.5)$$

Em (3.5),  $Y_t$  representa a sazonalidade, que são as oscilações de subidas ou quedas que ocorrem no período analisado. A componente sazonal possui movimentos facilmente previsíveis, ocorrendo em intervalos regulares de tempo, como por exemplo: Ano a ano, mês a mês ou mesmo dia a dia [15].

Enquanto  $a_t$  é um componente aleatório, de média zero e variância  $\sigma_a^2$ . Se  $a_t$  for um ruído branco, então  $E(a_t a_s) = 0$ ,  $s \neq t$  [12]. Quanto à variabilidade das observações, as séries temporais podem ser classificadas em:

- **Estacionárias:** Quando as estatísticas as suas estatísticas não são afetadas pela variação do tempo;
- **Ergódicas:** Quando apenas uma realização do processo estocástico é suficiente para se obter todas as estatísticas do mesmo.

A maioria dos métodos sobre de previsão baseia-se na ideia de que observações passadas contêm informações sobre o padrão de comportamento da série temporal. O propósito dos métodos é distinguir o padrão de qualquer ruído que possa estar contido nas observações e então usar esse padrão para prever valores futuros da série. Uma grande classe de modelos de previsão, tenta tratar ambas as causas de flutuações em séries de tempo e a das suavizações, médias móveis e amortecimento exponencial [12].

## 3.3 Modelagem de Séries Temporais

### 3.3.1 Modelos de Holt-Winters

O modelo de Holt-Winters descreve dados de demanda em que se verifica a ocorrência de tendência linear e componente de sazonalidade [13]. É muito utilizado para a previsão de curtíssimo prazo, devido a sua simplicidade, baixo custo de operação, boa precisão e capacidade de ajustamento automático e rápido às mudanças na série [14]. Este modelo possui duas equações de previsão, a aditiva e a multiplicativa. [12] sugerem que sugerem que o melhor critério de escolha entre os fatores multiplicativos ou aditivos é calcular

as formas multiplicativas e aditivas do algoritmo de Holt-Winters sazonal podem fornecer previsões bem diferentes. Se a série apresentar oscilações sazonais aproximadamente constantes, o modelo aditivo é mais indicado. Porém, se as oscilações sazonais forem proporcionais ao nível da série, o modelo multiplicativo é mais indicado. Alternativamente, pode-se utilizar os dois procedimentos e escolher aquele que fornece menor erro de previsão [16].

As duas formas do algoritmo de Holt-Winters sazonal (multiplicativas e aditivas) podem fornecer previsões bem diferentes. Se a série apresentar um modelo com oscilações sazonais aproximadamente constantes, o modelo aditivo é mais indicado. Porém, se as oscilações sazonais forem proporcionais ao nível da série, o modelo multiplicativo é mais indicado. Alternativamente, pode-se utilizar os dois procedimentos e escolher aquele que fornece menor erro de previsão [16].

Segundo [12], o modelo de Holt-Winters é baseado em três equações alisadoras, uma para o nível, outra para tendência e outra para sazonalidade. A sazonalidade pode ter efeito multiplicativo ou aditivo. As equações básicas para ambos os procedimentos do modelo são apresentadas na tabela 1.

Tabela 1: Equações Básicas dos modelos de Holt-Winters	
Modelo Multiplicativo	
Nível	$L_t = \alpha \frac{Y_t}{S_{t-2}} + (1 - \alpha) \times (L_{t-1} + B_{t-1})$
Tendência	$B_t = \beta(L_t - L_{t-1}) + (1 - \beta)B_{t-1}$
Sazonalidade	$S_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-2}$
Previsão	$F_{t+m} = (L_t + B_t + m)S_{t-2+m}$
Modelo Aditivo	
Nível	$L_t = \alpha(Y_t - S_{t-2}) + (1 - \alpha) \times (L_{t-1} + B_{t-1})$
Tendência	$B_t = \beta(L_t - L_{t-1}) + (1 - \beta)B_{t-1}$
Sazonalidade	$S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-2}$
Previsão	$F_{t+m} = L_t + B_t + m + S_{t-2+m}$

onde:

- $S$  - Comprimento de sazonalidade;
- $L_t$  - Nível da série;
- $B_t$  - Tendência;
- $S_t$  - Componente sazonal;
- $F_{t+m}$  - Previsão  $m$  passos à frente;

- $Y_t$  - Valor observado;
- $\alpha, \beta, \gamma$  - parâmetros de amortecimento do nível, da tendência e da sazonalidade, respectivamente.

A utilização do método de Holt-Winters, é bastante indicada para realizar previsões. Ela torna-se eficiente quando os dados de séries temporais apresentam tendência e sazonalidade.

### 3.3.2 Modelos de Box & Jenkins

Uma metodologia bastante utilizada na análise de modelos paramétricos é conhecida como abordagem de Box & Jenkins [2]. Tal metodologia consiste em ajustar modelos Auto-Regressivos integrados de médias móveis  $ARIMA(p, d, q)$  a um conjunto de dados. A construção é baseada em um ciclo iterativo, no qual a escolha da estrutura baseia-se nos próprios dados [12]. Segundo [15], a estratégia para a construção do modelo será feita com base em um ciclo iterativo ao qual a estrutura do modelo será escolhida baseada nos próprios dados.

As etapas do ciclo são:

- Especificação: Uma classe geral do modelo é considerada para análise;
- Identificação de um modelo, com base na análise de autocorrelações, autocorrelações parciais e outros critérios;
- Fase de estimação, na qual os parâmetros de modelo identificado são estimados;
- Fase de verificação ou diagnóstico do modelo ajustado, através de uma análise residual, para saber se este é adequado para fazer a previsão.

De acordo com [15], um processo estocástico pode ser definido como um modelo que descreve a estrutura de probabilidade de uma sequência de observações ao longo do tempo. Considerando a família  $Z = Z_t, t \in N$  tal que para cada  $t, Z_t$  é uma variável aleatória. Sendo  $Z_t$  originado de um experimento aleatório que pode ser repetido sob condições idênticas, a cada experimento obtém-se um registro dos valores de  $Z_t$  ao longo do tempo. Cada registro particular é uma realização do processo estocástico e uma série temporal é uma realização amostral do processo estocástico.

Um processo estocástico pode ser determinado quando são conhecidas as suas funções de distribuição probabilidade conjuntas, porém, como estas não são conhecidas e dispõe-se de apenas uma amostra do processo (a série temporal observada), pode assumir os pressupostos de estacionaridade e ergodicidade do processo estocástico [15].

A estacionaridade significa que as características do processo estocástico permanecem invariantes ao longo do tempo. Em um sentido estrito, a estacionaridade implica que as variáveis aleatórias  $Z_t$  e  $Z_{t+k}$  têm as mesmas distribuições de probabilidade para qualquer  $K$  escolhido.

Portanto, tendo como base como uma determinada série temporal gerada por um processo estocástico estacionário, onde o valor atual é dado por  $Z_t$ , Box & Jenkins propõem o modelo a seguir para descrever o processo estocástico gerador da série:

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}, \quad (3.6)$$

onde  $a_t$  é um ruído branco.

Segundo [15], na equação (3.6) as defasagens da variável  $Z_t$  no polinômio do lado esquerdo representam a parte autoregressiva do modelo ( $AR$ ), enquanto as defasagens dos choques aleatórios  $a_t$  no polinômio do lado direito representam a parte média móvel ( $MA$ ). A equação representa uma ampla parte dos modelos denominado  $ARMA(p, q)$ . Onde  $p$  representa a ordem de defasagem do termo de autoregressivo e o  $q$  a ordem a defasagem do termo de média móvel. Nota-se que, sendo  $p = 1$  e  $q = 0$  tem-se o termo autoregressivo de primeira ordem, ou seja  $AR(1)$ , no qual o valor da série no instante  $t$  depende somente do valor da série no tempo  $t - 1$ .

Para se aplicar a metodologia de Box & Jenkins a série em estudo deve ser pelo menos estacionaria de segunda ordem, ou seja, a série tem que ter média, variância, covariância finitas e constantes. O exame de estacionaridade pode ser realizado observando-se função de autocorrelação ou FAC da série temporal investigada. Em função das correlações entre os valores de uma série temporal [12]. O valor de  $Z_t$  depende probabilisticamente do valor de  $Z_{t-1}$ . Dessa forma, a previsão se torna possível por causa dessa dependência entre os valores da série é chamada de auto-correlação [15].

Um decaimento lento da FAC mostra que a série não que estacionária na média e precisa passar por uma ou mais diferenciações, se necessário. Além da FAC também é útil analisar o comportamento da Função de Autocorrelação Parcial (FACP), onde na medição da correlação entre duas observações seriais  $Z_{t+1}$  e  $Z_{t+k}$  elimina-se a dependência

dos termos intermediários  $Z_{t+1}, Z_{t+2}, \dots, Z_{t+k-1}$  [15].

A análise da FAC e da FACP é de fundamental importância para o procedimento de previsão de séries temporais, pois é com ela que são identifica-se as ordens  $p$  e  $q$  de um modelo *ARMA*.

No processo auto-regressivo *AR*, a FAC terá um decaimento exponencial ou em senoidal amortecida, se  $\phi_1 < 0$ , os sinais serão alternados. A FACP terá picos significativos nos lags 1 até  $p$ , depois cai a zero. Para um *AR* de ordem 1, o pico da FACP será no lag 1 depois cai a zero. O pico será positivo se  $\phi_1 < 0$  e negativo se  $\phi_1 < 0$ . No caso de um processo de médias móveis *MA*( $q$ ) a FAC terá picos nos lags 1 até  $q$  e depois cairá a zero. Em se tratando de um *MA*(1), o pico será no lag 1, caindo depois a zero. Caso se  $\phi_1 < 0$  será positivo o pico e se  $\phi_1 < 0$  será negativo. Na FACP há a presença de um decaimento exponencial ou senoidal amortecida. Pode-se verificar toda essa explicação de maneira resumida na tabela 2.

Tabela 2: Características principais dos modelos *AR*( $p$ ), *MA*( $q$ ) e *ARMA*( $p, q$ )

Características	<i>AR</i> ( $p$ )	<i>MA</i> ( $q$ )	<i>ARMA</i> ( $p, q$ )
Estrutura do Modelo	$\phi(B)Z_t = a_t$	$Z_t = \theta(B)a_t$	$\phi(B)Z_t = \theta(B)a_t$
FAC	Infinita	Finita	Infinita
FACP	Finita	Infinita	Infinita

Para [12], uma das características fundamentais da metodologia de Box & Jenkins é interpretar uma dada série temporal como sendo uma realização de um vetor aleatório multivariado, cuja dimensão é a da série temporal disponível. A partir de uma única realização do processo e, com os argumentos de estacionaridade e ergodicidade do processo subjacente, procura-se detectar o sistema gerador da série, através de informações contidas na mesma. A filosofia da modelagem de Box & Jenkins se utiliza de duas ideias: o princípio da parcimônia e a construção de modelos por meio de um ciclo iterativo. O princípio da parcimônia estabelece que deve-se escolher um modelo com o menor número possível de parâmetros, para uma adequada representação matemática. Um ciclo iterativo é uma estratégia de seleção de modelos a ser empreendida até que tenha-se um modelo satisfatório.

Se a série em estudo apresentar uma componente de tendência, então o processo estocástico gerador da série é não estacionário. Nesse caso a série deve passar por  $d$  diferenças simples para se tornar estacionária, condição básica para se aplicar a metodologia Box & Jenkins. Caso a primeira diferença não seja estacionária, o operador diferença deverá ser aplicado na série obtidas pelas diferenças simples e uma segunda filtragem

é efetuada, a qual poderá ser repetida quantas vezes necessárias até tornar-se a série estacionária [15].

A FAC e a FACP têm comportamentos teóricos conhecidos, entretanto na prática, como estas funções são estimadas, a análise dos gráficos da FAC e da FACP amostrais pode ser uma tarefa difícil, o que compromete a identificação precisa da ordem do modelo. A incorporação de coeficientes adicionais (modelos de maior ordem) melhora o grau de ajustamento, reduzindo a soma dos quadrados dos resíduos, no entanto, ressaltasse que modelos mais parcimoniosos produzem melhores previsões (maior capacidade de generalização) que os modelos sobre parametrizados.

Existem alguns critérios de seleção de modelos que permitem encontrar um ponto de equilíbrio entre a redução na soma do quadrado dos resíduos e a parcimônia do modelo. Os critérios mais usados são o AIC (Akaike Information Criterion) e o BIC (Bayesian Information Criterion), onde as formulas são dadas por:

$$AIC = T \ln(\hat{\sigma}_a^2) + 2n \quad (3.7)$$

$$BIC = T \ln(\hat{\sigma}_a^2) + n \cdot \ln(T), \quad (3.8)$$

onde  $n$  é o número de parâmetros estimados,  $T$  é o número de observações da série temporal e  $\hat{\sigma}_a^2$  é a estimativa de máxima verossimilhança da variância de  $a_t$  [12].

Se comparar os valores AIC e BIC de modelos com diferentes ordens, o melhor modelo é o que apresenta os menores valores nestas duas estatísticas, porém não deve-se basear apenas nestes dois critérios. A análise dos resíduos de modelos alternativos é de extrema importância na escolha do modelo final que melhor explica a dinâmica da série temporal em estudo [15]. A análise dos resíduos é feita através da FAC dos resíduos e do teste de Ljung-Box apresentado a seguir:

- $H_0$ : Os dados são independentemente distribuídos. Ou seja, todas as correlações são iguais a zero. ( $\rho_{i,j} = 0, \forall i \neq j$ );
- $H_1$ : Os dados não são independentemente distribuídos. Ou seja, existe correlação serial.

Uma vez que o  $p$  - valor for maior que o nível de significância (em geral, 0,05), então a hipótese nula  $H_0$  não é rejeitada e assim, os dados são independentes.

## 3.4 Filtragem SSA

Singular Spectrum Analysis (SSA), é uma técnica não paramétrica que permite decompor uma série temporal em sinal e ruído. É uma técnica útil para filtrar dados de séries temporais. [11] utilizou três metodologias na abordagem SSA: Análise de componentes principais (ACP), ACP associado com Análise de Cluster e Análise Gráfica dos Vetores Singulares. Em seu artigo é mostrado que o melhor método em SSA é a Análise Gráfica dos vetores singulares, que será usado neste projeto. SSA é um método recente e poderoso em séries temporais que incorpora elementos de análise clássica de séries temporais, estatística multivariada, geometria multivariada, sistemas dinâmicos e processamentos de sinais [4]. SSA tem sido aplicada com sucesso em diversas áreas: na matemática e física a economia e matemática financeira, na meteorologia e oceanografia a ciências sociais [8].

O método SSA é um procedimento que pode ser utilizado, dentre outras aplicações, na remoção de ruído e de séries temporais [8], [10]. A versão básica do método SSA pode ser dividida em duas etapas: decomposição e reconstrução.

### 3.4.1 Decomposição

Segundo [11], a etapa de decomposição pode ser subdividida em duas partes: Incorporação e decomposição em valores singulares (SVD – Singular Value Decomposition).

Seja  $Y_t = [y_1, \dots, y_T]_{1 \times T}$  uma série temporal [9] e considere  $L$  tal que  $2 \leq L \leq T$  de modo que  $L$  é um parâmetro a ser estimado e é chamado de comprimento da janela [8]. Entende-se por Incorporação o procedimento no qual uma série temporal  $Y_T$  é levada a uma matriz  $X = [y_1, \dots, y_k]_{L \times T}$ , para todo  $k \in [1, \dots, K]$ , onde  $K = T - L + 1$ . A matriz  $X$ , conhecida como matriz trajetória [10] é uma matriz Hankel, ou seja, os elementos de  $x_{i,j}$  tal que  $i + j = \text{constante}$  são iguais.

Considere  $S = XX'$ . Os autovalores de  $S$  dispostos em ordem de significância  $\lambda_1 \geq \dots \geq \lambda_L \geq 0$  são obtidos e os respectivos autovalores  $U_1, \dots, U_L$  são encontrados. Considere  $V' = (X'U_L)/\sqrt{\lambda}$ , como  $S$  é positivo semi-definido, então a matriz trajetória  $X$  pode ser expressa pela decomposição em valores singulares (SVD) apresentada em (3.9):

$$X = E_1 + E_2 + \dots + E_L, \quad (3.9)$$

onde  $E_l = \sqrt{\lambda}U_lV_l'$ , para todo  $l = 1, \dots, L$ . A coleção  $(\sqrt{\lambda_l}, U_l, V_l)$  é conhecida como autotripla da expansão SVD de  $X$ . Os elementos da autotripla são definidos respectiva-



mente por: valor singular, vetor singular à esquerda e vetor singular à direita de  $X$  [11]. A contribuição de cada componente em (3.9) pode ser mensurada pela razão de autovalores  $\lambda_l / \sum_{l=1}^L \lambda_l$ .

### 3.4.2 Reconstrução

Segundo [11], a etapa de reconstrução está subdividida em duas partes: 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.

Considere a sequência  $\sum_{l=1}^L E_l$  de matrizes elementares da expansão de SVD. Agrupe as mesmas em  $m$  grupos disjuntos utilizando algum método, por exemplo, por meio de análise de componentes principais, análise gráfica de vetores singulares ou agrupamento hierárquico e assumir que o conjunto de índices gerado é dado por  $\{I_1, \dots, I_m\}$ , de modo que a expansão (3.9) pode ser reescrita como em (3.10), sendo  $X_{I_i}$  arbitrária tal que  $X_{I_i} = \sum_{j=1}^{p_i} X_{I_{ij}}$  [11].

$$X = \sum_{l=1}^L E_l = \sum_{i=1}^m X_{I_i} \quad (3.10)$$

O objetivo do agrupamento é diminuir o número de componentes na expansão da matriz trajetória  $X$ . A contribuição de cada componente é mensurada pela razão (3.11) [11].

$$\frac{\sum_{j=1}^{p_i} \lambda_{I_{ij}}}{\sum_{l=1}^L \lambda_l}. \quad (3.11)$$

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

da matriz elementar  $X_{I_i}$  definida em (3.12), a partir da matriz elementar  $X_{I_i}$ .

$$y_t^{(i)} = \begin{cases} \frac{\sum_{l=1}^t x_{l,t-l+1}^{(i)}}{t}, & \text{se } 1 \leq t < L^* \\ \frac{\sum_{l=1}^{L^*} x_{l,t-l+1}^{(i)}}{L^*}, & \text{se } L^* \leq t < K^* \\ \frac{\sum_{l=t-K^*+1}^{T-K^*+1} x_{l,t-l+1}^{(i)}}{T-K^*+1}, & \text{se } K^* \leq t \leq T \end{cases} \quad (3.12)$$

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 autovalores  $\sum_{j=1}^{p_i} \lambda_{I_{ij}} / \sum_{l=1}^d \lambda_l$ . De acordo com [10], podemos classificar as componentes SSA  $\left[ y_t^{(i)} \right]_{1 \times T}$  de uma série temporal arbitrária  $[y_t]_{1 \times T}$  em três categoriais: *tendência*, *componentes harmônicas* (ciclo e sazonalidade) e *ruído* [8].

Um dos principais conceitos estudados em SSA é a propriedade de separabilidade [10]. Tal propriedade caracteriza quão bem separados estão as diferentes, componentes, umas das outras. Uma boa medida de separabilidade é a Correlação Ponderada. Por correlação ponderada *weighted correlation* ou *w*-correlação, podemos entender como uma função que quantifica a dependência linear entre duas componentes SSA  $Y_T^{(1)}$  e  $Y_T^{(2)}$  definida em (3.13) [11].

$$\rho_{ij}^{(w)} = \frac{\left( Y_T^{(i)}, Y_T^{(j)} \right)_w}{\|Y_T^{(i)}\|_w \|Y_T^{(j)}\|_w}. \quad (3.13)$$

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

Através da separabilidade, pode-se verificar estatisticamente se duas componentes SSA estão bem separadas, em termos de dependência linear. Se o valor absoluto da *w*-correlação é pequeno [10], 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 SSA [8].

## 3.5 Resumo da Metodologia

No Capítulo 2, foram expostos os objetivos deste trabalho. A figura 1, mostra exatamente o que será feito para efeito de comparação. Será introduzida uma série onde a mesma passará pela filtragem SSA, onde será dividida em 3 partes: Tendência, Harmônica e Ruído. Após a filtragem será retirado o Ruído, onde a nova série apenas com as partes: Tendência e Harmônica será analisada e feita a comparação com a mesma série sem a filtragem SSA. A partir das duas séries: original e filtrada será feita a modelagem ARIMA e de Holt-Winters para análises.

Os softwares utilizados para o estudo foram: FPW (Forecast Pro for Windows) para fazer a modelagem Holt-Winters e Box & Jenkins e CaterpillarSSA para abordagem SSA via análise gráfica dos componentes principais. Os resultados obtidos ao longo dos experimentos computacionais realizados são comparados em termos das estatísticas de aderência: MAPE (Mean Absolute Percentage Error) (3.14), RMSE (Root Mean Squared Error) (3.15), BIC (Bayesian Information Criterion) (3.8) e  $R^2$  (coeficiente de determinação) (3.16).

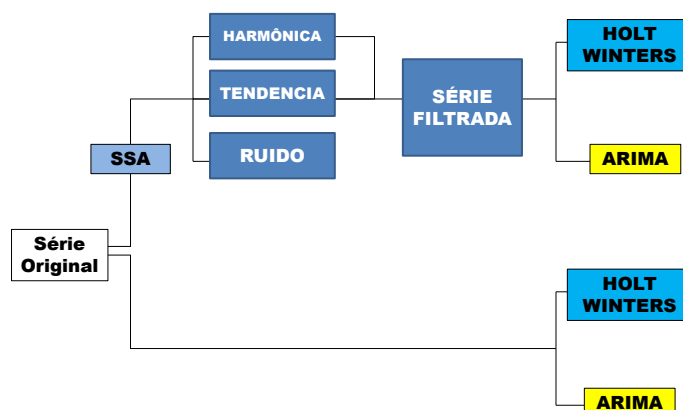


Figura 1: Fluxograma da metodologia.

$$MAPE = \sum_{t=1}^T \left[ \frac{|y_t - \hat{y}_t|}{y_t} \right] \times \frac{1}{T} \quad (3.14)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{T}} \quad (3.15)$$

$$R^2 = 1 - \left( \frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{\sum_{t=1}^T (y_t - \bar{y})^2} \right) \quad (3.16)$$

## 4 Análise dos Resultados

### 4.1 Filtragem SSA

A figura 2 apresenta série original, medida a cada 10 minutos, em um intervalo de uma semana, contemplando um total de 1008 observações.

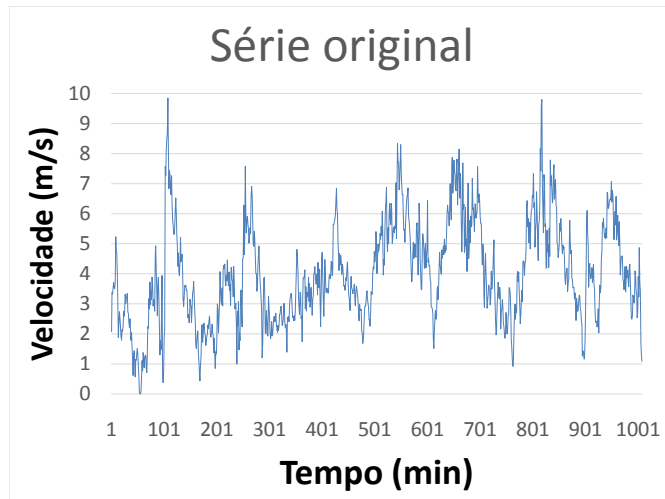


Figura 2: Série original de velocidade do vento medida a cada 10 minutos.

Com o objetivo de realizar filtragem SSA da série original, foi considerado o comprimento de janela  $L = T/2 = 504$ . Os autovetores (ou vetores singulares) foram obtidos na fase de decomposição da série. A figura 3 apresenta os 9 primeiros vetores singulares e suas respectivas contribuições para a série. É possível verificar que os vetores 1 e 4 pertencem a componente de tendência e que os vetores 2 e 3 pertencem a componente harmônica. Os demais vetores necessitam de uma análise mais profunda. Essa análise será feita pela análise do periodograma, que é a análise gráfica das componentes em pares seguidos.

A figura 4 apresenta alguns gráficos de dispersão dos vetores singulares em sequência. Pode-se ver que os pares de vetores (1 - 2), e (10 - 11) e (33 - 34) não apresentam comportamento que nos dê alguma informação. Isso significa que os vetores em cada par

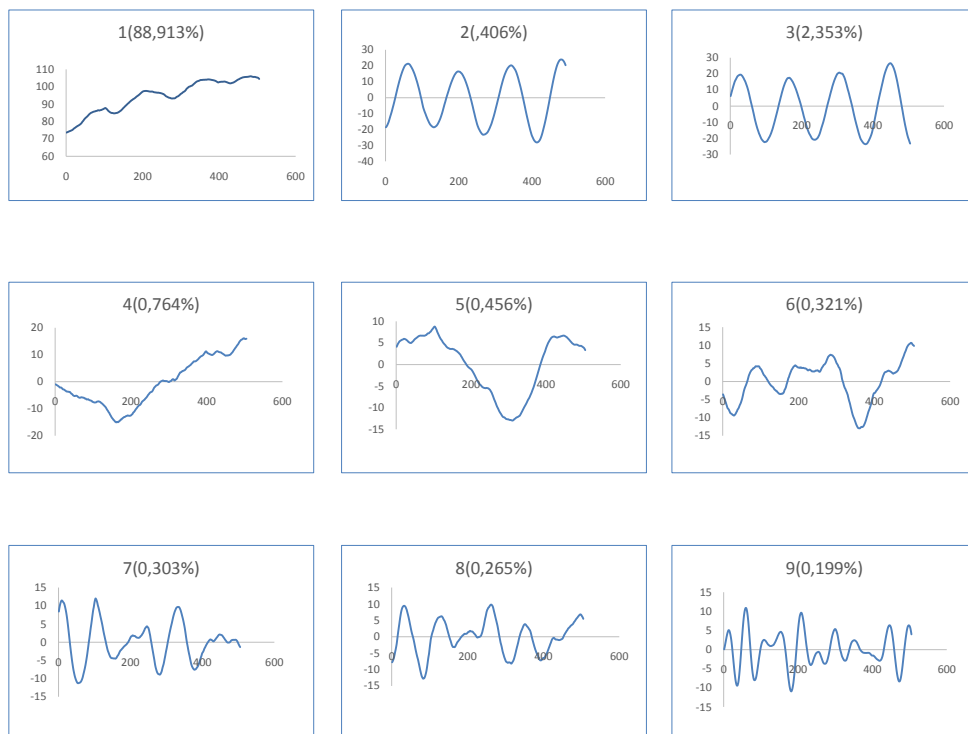


Figura 3: Os 9 primeiros vetores singulares e suas respectivas contribuições à série.

não pertencem a mesma componente, como é o caso, por exemplo, do par (1 - 2) onde o vetor 1 pertence a componente de tendência e o vetor 2 pertence a componente harmônica. Os vetores (2 - 3) pertencem a componente harmônica, pois formam um círculo ou um polígono regular e seu período e o número de lados do polígono. Já os vetores (4 - 5) pertencem a componente de tendência. Por sua vez, os vetores (434 - 435) pertencem a componente ruidosa e seu gráfico de dispersão mostra exatamente algo totalmente caótico.

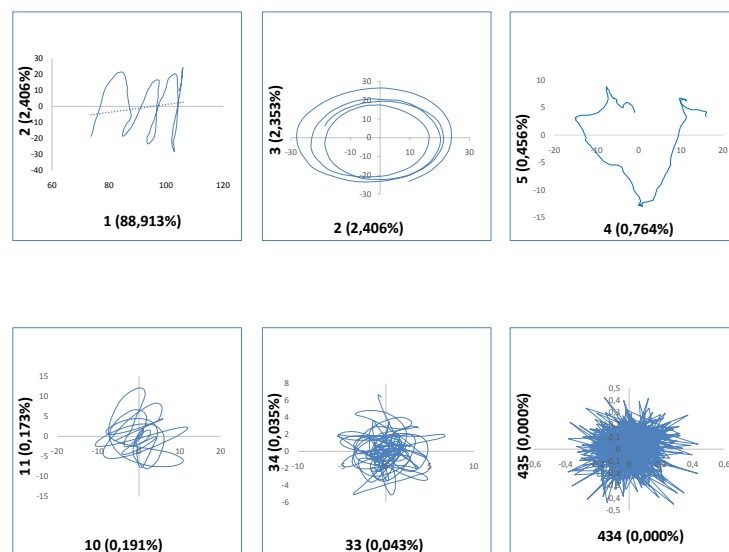


Figura 4: Gráfico de dispersão de 6 pares dos autovetores.

Após as análises gráficas, os vetores singulares foram alocados conforme suas classes de componentes. A tabela 3, mostra exatamente como eles foram agrupados.

Tabela 3: Agrupamento dos Vetores Singulares após a Análise Gráfica.

Componente	Vetores Singulares
Tendência	1, 4 – 6, 8, 11
Harmônica	2, 3, 7, 9, 10, 12 – 33, 36 – 43, 48, 49
Ruído	34, 35, 44 – 47, 50 – 504

Feito o agrupamento dos vetores singulares, a série está bem dividida em: Tendência, Harmônica e Ruído. Na figura 5, segue o gráfico de cada componente de maneira separada, como pode-se notar graficamente as componentes estão bem definidas.

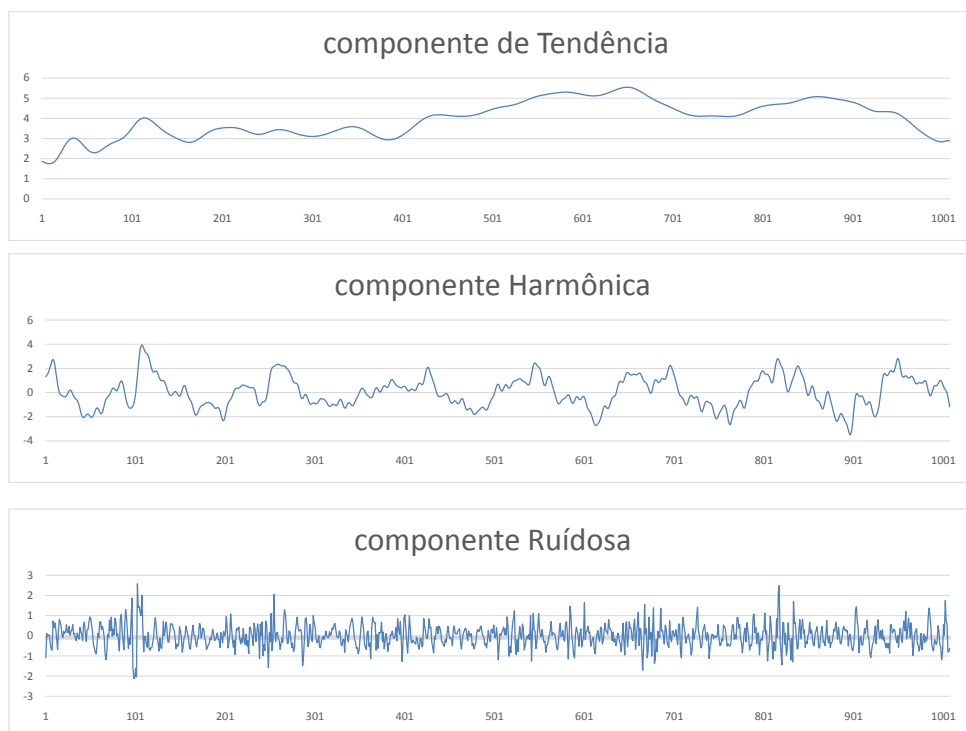


Figura 5: As três Componentes obtidas no Agrupamento SSA.

Outra maneira de verificar se as componentes estão bem separadas é pela suas correlações, ou seja, quanto menor a correlação entre cada componente melhor foi a separação, significa que tem o mínimo de alocação dos vetores singulares em componentes erradas. A tabela 4 mostra a matriz de correlação ponderada das três componentes, utilizando a análise gráfica dos vetores singulares. Segundo os valores apresentados na tabela, as componentes obtiveram uma estrutura de separabilidade bem definida.

Tabela 4: Correlação Ponderada entre as três Componentes SSA.

Componente	Tendência	Harmônica	Ruído
Tendência	1	0,02	0,05
Harmônica	0,02	1	0,001
Ruído	0,001	0,05	1

Como a correlação foi muito pequena, pode-se dizer que cada componente está bem definida, ou seja, os vetores singulares estão bem classificados. Por fim, a série filtrada via SSA seguirá apenas com as componentes de tendência e harmônica. A componente ruído foi retirada, logo têm-se duas séries a série original e a serie filtrada, sem o ruído. Como o objetivo do trabalho é comparar as duas séries: Original e filtrada. Segue, na figura 6, a série original sendo sobreposta pela série filtrada após a abordagem SSA através análise gráfica dos vetores singulares. Pela figura 6 nota-se que a série filtrada apresenta um comportamento mais suave que a série original, mostrando também que a separação foi bem realizada e que o ruído foi extraído.

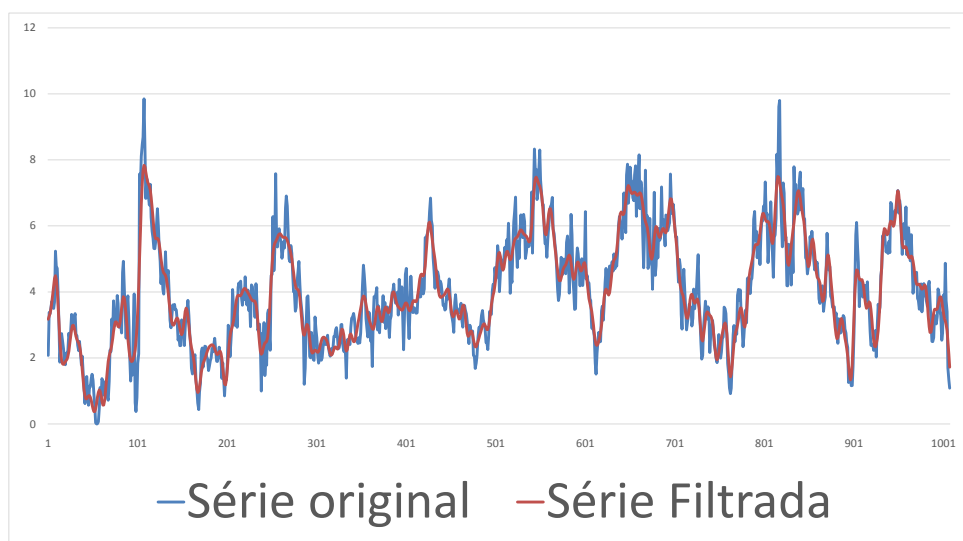


Figura 6: Comparação gráfica entre a Série Original e a Série Filtrada SSA.

## 4.2 Modelagem

Após fazer as análises gráficas, e assim fazer a separação das componentes e fazer a filtragem da série original, originando uma nova série, composta apenas pelas compo-



mentes tendência e harmônica, é necessário comparar os modelos de Holt-Winters e de Box & Jenkins via estatísticas de aderência. Já foi visto que separação foi bem feita, visto pela correlação ponderada, e também que a série filtrada tem um gráfico com um comportamento mais suave, ou seja, sem ruído. As estatísticas de aderência são feitas para modo de comparação do melhor modelo. Nesse projeto será utilizado as seguintes estatísticas: MAPE, RMSE, BIC e  $R^2$ . Para essas estatísticas, a comparação é feita pelo menor valor, ou seja, quanto menor for o valor da estatística melhor é o modelo em termos de capacidade preditiva. Já o  $R^2$  acontece de forma inversa, quanto maior for o  $R^2$ , mais bem explicada está a série pelo modelo.

Para a verificação do poder preditivo com o uso da abordagem SSA na série temporal da série de velocidade do vento medida a 50 metros com periodicidade de 10 em 10 minutos, foram utilizados os métodos preditivos de Holt-Winters e de Box & Jenkins. Na tabela 5 encontram-se os resultados das estatísticas de aderência.

Na tabela 5, pode-se verificar que o poder preditivo dos modelos quando usados a série filtrada via SSA supera em muito o poder preditivo dos modelos quando aplicados na série original. Essa verificação é medida pelos erros, que são todos menores para a série filtrada. Em relação ao  $R^2$ , este é maior para os modelos aplicados a série filtrada mostrando, assim, que a série filtrada é muito melhor para a previsão que a série original.

Tabela 5: Estatísticas de Aderência aplicada aos Modelos Testados.

Série	Modelo	MAPE	RMSE	BIC	$R^2$
Original	Holt-Winters	0,1987	0,6704	0,6727	0,8431
Original	Box & Jenkins	0,2566	0,6614	0,6659	0,8473
Filtrada SSA	Holt-Winters	<b>0,0209</b>	<b>0,0838</b>	<b>0,0844</b>	<b>0,9972</b>
Filtrada SSA	Box & Jenkins	<b>0,0053</b>	<b>0,0218</b>	<b>0,0225</b>	<b>0,9997</b>

Visto que a série filtrada via SSA proporcionou uma melhora significativa comparada a série original quando modelada, outro fator importante que também pode ser verificado na tabela 5 é que a modelagem de Box & Jenkins para a série filtrada teve melhor ajustes, ou seja, os erros de previsão foram menores.

### 4.3 Previsão

Para consolidar estes resultados os dados foram divididos em duas partes: Dentro da amostra (in sample) e fora da amostra (out of sample), onde a primeira parte são utilizados 912 das 1008 observações com fase de treinamento do modelo e na segunda

parte, 96 observações finais como fase de teste do modelo. Na primeira parte, os valores modelados são obtidos a partir dos dados históricos da série. Na segunda parte, os valores modelados são obtidos a partir dos valores anteriores do próprio modelo, configurando-se assim, como previsões que podem ser comparadas com valores reais podendo, desta forma, comparar os resultados tanto na modelagem como na previsão. Na tabela 6 estão os resultados das estatísticas de aderência dentro da amostra e na tabela 7, estes resultados fora da amostra.

Tabela 6: Estatísticas de Aderência aplicada aos Modelos Testados (In Sample).

Série	Modelo	MAPE	RMSE	BIC	$R^2$
Original	Holt-Winters	0,2057	0,6793	0,6818	0,8437
Original	Box & Jenkins	0,3809	0,6689	0,6739	0,8484
Filtrada SSA	Holt-Winters	<b>0,0209</b>	<b>0,0821</b>	<b>0,0827</b>	<b>0,9974</b>
Filtrada SSA	Box & Jenkins	<b>0,0030</b>	<b>0,0116</b>	<b>0,0121</b>	<b>0,9999</b>

Tabela 7: Estatísticas de Aderência aplicada aos Modelos Testados (Out of Sample).

Série	Modelo	MAPE	RMSE	MAD
Original	Holt-Winters	0,123	0,988	0,452
Original	Box & Jenkins	0,1259	1,087	0,459
Filtrada SSA	Holt-Winters	<b>0,021</b>	<b>0,638</b>	<b>0,083</b>
Filtrada SSA	Box & Jenkins	<b>0,005</b>	<b>0,085</b>	<b>0,017</b>

Pode se verificar que os resultados in sample e out of sample continuam com valores melhores em todas as estatísticas de aderência para a série filtrada. Os modelos testados foram escolhidas pelo próprio programa FPW como sendo os modelos mais adequados para as séries original e filtrada, segundo os testes de normalidade e estacionaridade.

## 5 Conclusão

A energia de origem hídrica passa por momentos de estiagens e nesses momentos existe a necessidade de uma energia complementar, onde a mesma não influa drasticamente os recursos naturais. Assim, uma energia limpa e renovável foi colocada à disposição para complementar a energia hidrelétrica, que foi a energia eólica. Visto também em artigos, que existe uma correlação negativa entre a energia eólica com a hidrelétrica, ou seja, em tempos de estiagem notou-se um aumento dos ventos e suas velocidades.

Portanto, verificou-se a necessidade de encontrar um modelo para fazer previsões da velocidade do vento em um determinado espaço de tempo, assim no tópico materiais e métodos foi exposto os métodos utilizados para buscar esse melhor modelo de previsão. Foi introduzido também uma ferramenta fundamental para essa modelagem que permite a retirada do ruído da série: a abordagem SSA. Com isso é comparado a série original com a série filtrada e verifica-se, segundo as estatísticas de aderência, a melhor série para modelar a velocidade do vento.

É importante destacar que o objetivo desse projeto foi mostrar que os modelos aplicados à série filtrada detém melhores capacidades preditivas que os modelos aplicados à série original. Desat forma, foi feita a filtragem da série original obtendo uma série menos ruidosa e a partir desta e da série original foram feitas modelagens de Holt-Winters e de Box & Jenkins. O processo de filtragem SSA foi feito com a separação das componentes da serie original: Tendência, Harmônica e Ruído. A separação foi bem feita, segundo a correlação ponderada entre as componentes. Feito isso, o ruído da série foi removido, tornando-se uma nova série que foi chamada “Série Filtrada”. Os modelos de Holt-Winters e de Box & Jenkins foram aplicados as duas séries e foram comparados segundo as estatísticas de aderência baseadas na diferença entre os valores observados e os valores modelados.

Os resultados demostram que série filtrada obteve valores muito melhores que a série original, quando modeladas para previsão. Após essa verificação, foi também ressaltado a

análise in sample e out of sample, ou seja a previsão dentro da amostra e fora da amostra, respectivamente, que da mesma forma, os valores continuaram melhores. Desta forma, podemos concluir que realizar a filtragem na série temporal antes de modelá-la incorre em ganho preditivo e que é possível utilizar a energia eólica como fonte alternativa à energia hidrelétrica nos momentos de estiagem e que o melhor modelo para previsão da velocidade do vento de curtíssimo prazo é o modelo  $ARIMA(p, d, q)$  com transformação logarítmica para a série filtrada via SSA.

## Referências

- [1] Amarante, O. A. C., Brower, M., Cack, J. Sá, A. L. *Atlas do Potencial Eólico Brasileiro*. Brasília, 2001.
- [2] Box, G. E. P., and Jenkins, G. M. *Time series analysis: Forecasting and control*. Holden-day, 1970.
- [3] Dalmaz, A. *Estudo do Potencial Eólico e Previsão de ventos Para Geração de Eletricidade em Santa Catarina*, Dissertação de Mestrado, UFSC, Florianópolis, 2007.
- [4] Elsner, J. B. and Tsonis, A. A. *Singular Spectrum Analysis. A New Tool in Time Series Analysis*. Plenum Press, 1996.
- [5] Frasso, Lucas , Barra, Menicone. *Entenda como a geração de energia elétrica afeta o meio ambiente*, São Paulo, 23 março de 2011. Saúde e Ciência. Disponível em: <http://g1.globo.com/ciencia-e-saude/noticia/2011/03/entenda-como-geracao-de-energia-eletrica-afeta-o-meio-ambiente.html>.
- [6] Garbe, E. A., Mello, R., Tomaselli, I., "Projeto conceitual e Análise de viabilidade Econômica de Unidade de Geração de Energia Elétrica Eólica na Lagoa dos Patos - RJ", *Economia & Energia*, Ano: XV-No 83 outubro/dezembro, 2011.
- [7] Goldemberg, J. & Villanueva, L. D. *Energia, Meio Ambiente e Desenvolvimento*, Edusp São Paulo. Vol I, 1-1, 2003.
- [8] Golynadina, N., Nekrutkin, V., and Zhihgljavsky, A. *Analysis of time series structure: SSA and reletade techniques*. Chapman&Hall/CRC, New York, USA, 2001.
- [9] Hamilton, J. *Time Series Analysis*. Princeton University Press, 1994.
- [10] Hassani, H., Heravi, S., and. Zhigljavsky, A. "Forecasting UK Industrial Production with Multivariate Singular Spectrum Analysis," presented at the 2012 International Conference on the Singular Spectrum Analysis and its Applications, Beijing, China, 2012.
- [11] Menezes, M. L, Cassiano, K. M., Souza, R. M., Junior, L. A. T., Pessanha, J. F. M. e Souza, R. C. Modelagem e previsão de demanda de energia com filtragem SSA. *Revista da Estatística UFOP*, 3 (2), 2004.
- [12] Morettin, P. A., Toloi. *Análise de Séries Temporais*, 2a. edição. Blucher, São Paulo, 2006.
- [13] Pellegrini, F. R., Fogliatto, F. Estudo comparativo entre modelos Winters e Box Jenkins para a previsão de demanda sazonal. *Revista Produto & Produção*. V.4, número especial, 2000.

- [14] Serra, C. M. V. Aplicação de séries temporais na análise de demanda turística no estado do Pará usando os modelos de Holt-Winters, In XXV Encontro Nacional de Engenharia de Produção, Porto Alegre, 2005.
- [15] Silva, I. V. G. A previsão do vento para geração de energia eólica. Dissertação de mestrado, PUC, departamento de engenharia elétrica. Rio de Janeiro, 2010
- [16] Vasconcelos, A. S., Costa, J. H. F., Análise de modelos de séries temporárias para a previsão mensal do ICMS do Estado do Maranhão, São Luís: IMESC, 2008.