



Larissa Moraes Dantas Campos

**Previsão de velocidade do vento utilizando
*Singular Spectrum Analysis***

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Engenharia de Produção do Departamento de Engenharia Industrial da PUC-Rio.

Orientador: Prof. Fernando Luiz Cyrino Oliveira
Co-orientadora: Prof. Paula Medina Maçaira Louro



Larissa Moraes Dantas Campos

**Previsão de velocidade do vento utilizando
*Singular Spectrum Analysis***

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Engenharia de Produção da PUC-Rio. Aprovada pela Comissão Examinadora abaixo.

Prof. Fernando Luiz Cyrino Oliveira

Orientador

Departamento de Engenharia Industrial - PUC-Rio

Prof. Paula Medina Maçaira Louro

Co-Orientadora

Departamento de Engenharia Industrial - PUC-Rio

Prof. Reinaldo Castro Souza

Departamento de Engenharia Industrial - PUC-Rio

Prof. Soraida Aguilar Vargas

Departamento de Engenharia Industrial – PUC-Rio

Rio de Janeiro, 29 de abril de 2020

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, da autora e do orientador.

Larissa Moraes Dantas Campos

Graduou-se em Engenharia de Produção pela Universidade do Estado do Pará – UEPA, em 2018, atua na área de pesquisa de séries temporais de vento.

Ficha Catalográfica

Campos, Larissa Moraes Dantas

Previsão de velocidade do vento utilizando Singular Spectrum Analysis / Larissa Moraes Dantas Campos ; orientador: Fernando Luiz Cyrino Oliveira ; co-orientadora: Paula Medina Maçaira Louro. – 2020.
68 f. : il. color. ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Industrial, 2020.
Inclui bibliografia

1. Engenharia Industrial - Teses. 2. Séries temporais de vento. 3. SSA. 4. Previsão. 5. Curto prazo. 6. Brasil. I. Oliveira, Fernando Luiz Cyrino. II. Louro, Paula Medina Maçaira. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Industrial. IV. Título.

CDD: 658.5

Dedico este trabalho ao meu marido, Dionísio, minha mãe Ana Lúcia, meu pai Kleber e meu irmão Kleber Filho, razões de eu ter chegado até aqui.

Agradecimentos

Agradeço primeiramente a Deus que me deu forças, me sustentou e é meu maior apoio, me ajudando a superar os dias difíceis e não desistir diante das adversidades.

Agradeço ao meu marido, Dionísio, meu grande exemplo de persistência e determinação, maior incentivador dos meus sonhos, por estar ao meu lado durante essa jornada nos momentos em que pensei em desistir, por me compreender e me amar, me dando condições de seguir em frente quando tudo parecia impossível.

Aos meus pais, Ana Lúcia e Kleber, por sempre serem meu suporte, emocional e financeiro, buscarem sempre dar o seu melhor para que eu chegasse onde estou hoje, são a razão de eu buscar ser melhor a cada dia. E ao meu irmão, Kleber Filho, por todas as conversas e amizade que foram fundamentais durante esse período.

Agradeço também aos meus sogros, Maria do Socorro e Dionísio Filho, e a minha cunhada Suelen Karyna por nunca deixarem de interceder por mim, e mesmo com a distância sempre se fazerem presentes.

Aos meus pastores Bruno Rafael e Denise, por todo apoio e orações durante os momentos difíceis nessa fase, intercedendo e me acompanhando dia após dia nessa jornada.

Aos meus amigos do Pará, que torceram e oraram por mim durante esse período, obrigada por toda mensagem de apoio, todas as brincadeiras e risos, em especial meus amigos da UEPA (o G10), que sempre me deram forças para continuar.

Agradeço também aos amigos que fiz no Rio de Janeiro, Brenda, Amanda, Híngred, onde juntos compartilhamos alegrias, aflições, e trilhamos juntos esse caminho, além de me acolherem em suas casas ou me fazerem companhia quando me sentia sozinha, à Margareth por todos os conselhos e apoio nos momentos de aflição.

Em especial, agradeço a Marla Catarine e William Rosseto, onde juntos foram minha família durante esse período árduo, estavam ao meu lado em todos os momentos me ajudando a superar vários obstáculos, sendo meu apoio e me ajudando a passar pelos dias difíceis, trazendo conforto com suas palavras e me incentivando a prosseguir. Obrigada por essa amizade que vou levar pra vida!

Agradeço a professora Mariana Carneiro por me incentivar a seguir em minha carreira e me auxiliar em diversos momentos durante esse período.

Agradeço a CAPES pelo auxílio financeiro, a Puc-Rio pela oportunidade de fazer parte desse programa, e a toda equipe do DEI, em especial, Dona Graça, por todos os cafés da tarde que animavam meu dia, por todo cuidado e carinho durante esses dois anos.

Agradeço também aos meus orientadores Fernando Cyrino e Paula Maçaira por todo conhecimento compartilhado para a realização dessa pesquisa.

Por fim, a todos que contribuíram direto ou indiretamente para minha formação e me ajudaram a chegar até aqui, meus agradecimentos.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

Campos, Larissa Moraes Dantas; Cyrino Oliveira, Fernando Luiz (Orientador); Louro, Paula Medina Maçaira (Coorientadora). **Previsão de velocidade do vento utilizando Singular Spectrum Analysis**. Rio de Janeiro, 2020. 68p. Dissertação de Mestrado – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

Uma mudança de paradigma no mundo todo foi ocasionada pelo aumento da preocupação quanto ao uso de combustíveis fósseis usados como principal fonte de geração elétrica, a correspondente mudança climática e os danos ambientais crescentes. Nos últimos anos, a energia eólica apresentou um crescimento incessante como alternativa sustentável para a produção de eletricidade, o que pode ser observado a partir do crescimento de sua capacidade instalada mundialmente. O Brasil está entre os dez países que tem as maiores capacidades instaladas, e apresentou 9,42% de geração de energia elétrica advinda da fonte eólica em 2019. No entanto, a aleatoriedade e a intermitência do vento são os maiores desafios na integração dessa fonte no sistema de energia. Diante deste contexto, esta pesquisa propõe a aplicação da técnica *Singular Spectrum Analysis* (SSA) como método de previsão para uma série de velocidade eólica no Brasil, fazendo uma análise comparativa de modelos SSA considerando diferentes horizontes de previsão e conjunto de treinamento para diferentes dias de previsão, com diferentes tamanhos de série temporal. Deste modo, é comparada a série temporal do ano todo com somente o último mês desta série para prever os últimos sete dias do mês de dezembro. Os resultados dessa aplicação mostram que para a maioria dos dias a utilização do ano todo como conjunto de treinamento obteve melhor desempenho, indicando que o uso da técnica SSA pode ser uma alternativa para séries temporais com uma grande quantidade de dados.

Palavras-chave

Séries temporais de vento; SSA; previsão; curto prazo; Brasil.

Abstract

Campos, Larissa Moraes Dantas; Cyrino Oliveira, Fernando Luiz (Advisor); Louro, Paula Medina Maçaira (Coadvisor). **Wind speed prediction using Singular Spectrum Analysis**. Rio de Janeiro, 2020. 68p. Dissertação de Mestrado – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

A paradigm shift around the world was caused by increased concern about the use of fossil fuels used as the main source of electricity generation, the corresponding climate change and increasing environmental damage. In recent years, wind energy has shown steady growth as a sustainable alternative for electricity production, which can be seen from the growth of its installed capacity worldwide. Brazil is among the ten countries that have the largest installed capacities, and presented 9.42% of electricity generation from the wind source in the last year. However, wind randomness and intermittency are the biggest challenges in integrating this source into the energy system. In this context, this research proposes the application of the Singular Spectrum Analysis (SSA) technique as a forecast method for a series of wind speed in Brazil, making a comparative analysis of SSA models considering different forecast horizons and training set for different days forecast, with different time series sizes. In this way, the time series of the whole year is compared with only the last month of this series to forecast the last seven days of the month of December. The results of this application show that for most days the use of the whole year as a training set obtained better performance, indicating that the use of the SSA technique can be an alternative for time series with a large amount of data.

Keywords

Time series of wind; SSA; forecast; short term; Brazil.

Sumário

1	Introdução	13
2	Referencial Teórico	20
2.1	Previsão de Séries Temporais	20
2.2	Previsão da velocidade do vento de curto prazo	22
2.3	<i>Singular Spectrum Analysis (SSA)</i>	25
3	<i>Singular Spectrum Analysis</i>	28
3.1	1º Estágio: Decomposição	29
3.1.1	1ª Etapa: Incorporação	29
3.1.2	2ª etapa: Decomposição em Valor Singular (SVD)	30
3.2	2º Estágio: Reconstrução	31
3.2.1	3ª etapa: Agrupamento	31
3.2.2	4ª etapa: Média diagonal	32
3.3	Previsão com SSA	33
3.4	Informações Adicionais	34
3.4.1	Tamanho de janela	34
3.4.2	Escolha das autotriplas	35
3.4.3	Valores singulares	36
3.4.4	Correlação ponderada	37
4	Resultados	39
4.1	Descrição da base de dados	39
4.2	Aplicações	46
4.2.1	Primeira aplicação	46
4.2.2	Segunda aplicação	53
4.3	Comparação entre as aplicações	60
5	Conclusões e pesquisas futuras	64
	Referências bibliográficas	66

Lista de Figuras

Figura 1 - Crescimento da Capacidade Eólica Instalada no Mundo	13
Figura 2 - Crescimento da capacidade eólica instalada no Brasil	14
Figura 3 - Matriz Elétrica Nacional.	15
Figura 4 - Relação entre geração hidrelétrica e geração eólica	16
Figura 5 - Geração de Energia Eólica (GWh)	16
Figura 6 - Velocidade do vento no decorrer do dia (em m/s)	17
Figura 7 - Etapas da técnica SSA	29
Figura 8 - Exemplo de gráfico de valores singulares	37
Figura 9 - Exemplo de matriz de correlações – w	38
Figura 10 - Série temporal com as médias das velocidades diárias.	40
Figura 11 - <i>Boxplot</i> das observações	41
Figura 12 - <i>Boxplot</i> com os dados de cada mês	41
Figura 13 - Histograma com a frequência das velocidades	42
Figura 14 - Fluxograma com as etapas para realização da pesquisa	44
Figura 15 - Conjuntos de treinamentos para os dias de previsão	47
Figura 16 - Gráfico dos Autovalores	49
Figura 17 - Matriz de Correlações	50
Figura 18 - Gráfico de autovetores pareados	51
Figura 19 - Previsões de 24 horas para os tamanhos de janela para o dia 31/12	52
Figura 20 - Conjunto de treinamento para o dia previsto	54
Figura 21 - Gráfico de autovalores para $L = 120$	56
Figura 22 - Matriz de correlações para $L = 120$	56
Figura 23 - Gráfico de autovetores pareados para $L = 120$	57
Figura 24 - Previsões 24 horas para os tamanhos de janela para o dia 31/12 na segunda aplicação	58
Figura 25 - Comparação entre o desempenho das previsões com RMSE	61

Figura 26 - Comparação entre o desempenho das previsões com SMAPE

62

Figura 27 - Variação do RMSE no dia 31/12

63

Figura 28 - Variação do RMSE no dia 25/12

63

Lista de Tabelas

Tabela 1 - Estatísticas da série temporal de velocidade do vento	40
Tabela 2 - Valores de L e r testados	48
Tabela 3 - Métricas para cada par de parâmetro testado	52
Tabela 4 - Melhor par de parâmetros de acordo com o dia previsto	53
Tabela 5 - Valores de L e r testados para a segunda aplicação	55
Tabela 6 - Métricas para os parâmetros na segunda aplicação	59
Tabela 7 - Melhor par de parâmetros de acordo com o dia previsto	60
Tabela 8 - Valores do RMSE para as duas aplicações	60
Tabela 9 - Valores do SMAPE para as duas aplicações	61

1

Introdução

A crescente preocupação quanto ao uso de combustíveis fósseis como meio de geração de energia, a crise associada a eles e a correspondente mudança climática têm levado a uma mudança de paradigma em todo o mundo. Fontes alternativas de energia verde, que são não esgotáveis, sustentáveis e têm baixa emissão de carbono são mais desejadas para a geração de energia (Aasim *et al.*, 2019). A energia eólica é uma das fontes mais disponíveis, acessíveis e eficientes (Chen & Folly, 2018), além de ser econômica e ecologicamente correta, o que a torna uma das alternativas mais populares e promissoras (Zhao *et al.*, 2016).

Nos últimos anos, a energia proveniente dos ventos apresentou um crescimento incessante como alternativa sustentável para a produção de eletricidade (López-Manrique *et al.*, 2018). Esse crescimento pode ser observado em todo o mundo a partir do aumento da capacidade instalada no decorrer dos anos, como pode ser visto na Figura 1, a partir de dados obtidos pela *Global Wind Energy Council* – GWEC, para o ano de 2017 (GWEC, 2018).



Figura 1 - Crescimento da Capacidade Eólica Instalada no Mundo

Fonte: Adaptado de GWEC (2018)

O Brasil, teve participação de cerca de 4% dessa capacidade no ano de 2017, com 2.022 MW, e está entre os dez que possuem maiores capacidades instaladas no mundo, e em relação a América Latina e Caribe, sendo o primeiro no *ranking* no mesmo ano (GWEC, 2018).

No país, esse crescimento pode ser notado pelo aumento da capacidade instalada no decorrer do tempo, pelo crescimento do número de instalações novas todos os anos, além de projeções positivas para os anos seguintes (Figura 2) (ABEEÓLICA, 2019). O Brasil terminou 2019 com a marca de 15,4 GW de capacidade instalada, além de mais 2 GW de capacidade em construção, com 619 usinas instaladas (ABEEÓLICA, 2020). O país começou a prestar atenção ao potencial da energia eólica em 2001, em função da crise energética que aconteceu no país, chamada de “apagão”. Deste modo, foi preciso diversificar a matriz elétrica e a geração eólica em uma alternativa de rápida implementação (Menezes, 2019).

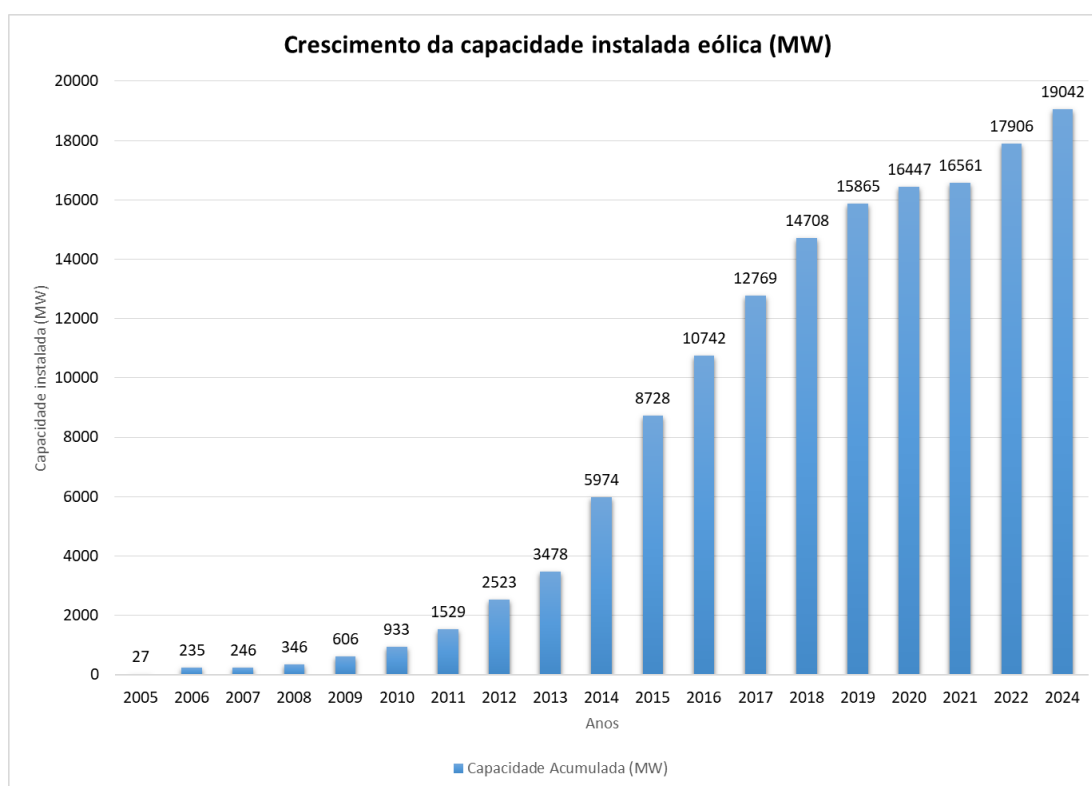


Figura 2 - Crescimento da capacidade eólica instalada no Brasil

Fonte: Adaptado de ABEEÓLICA (2019)

De acordo com a matriz elétrica brasileira (Figura 3) observa-se que dentre as diversas fontes que contribuem para a geração de eletricidade a mais utilizada ainda para produção de eletricidade no Brasil até então é a hidrelétrica, gerando 70,45% no último ano a partir dela (ONS, 2020). Ela tem sido a principal fonte de geração do sistema do país durante várias décadas, tanto pela abundância do recurso em todo o país, como também pela sua competitividade econômica (EMPRESA DE PESQUISA ENERGÉTICA - EPE, 2017). Mas, complementar a

ela, a presença de usinas termelétricas é vista como um fator estratégico no setor elétrico, visto que, devido a relevância da participação das fontes hídricas para a geração de energia no Brasil, as fontes térmicas têm atuado em períodos de escassez hidrológicas (EPE, 2017), por isso ela se encontra na segunda posição como fonte mais utilizada para a geração de energia.

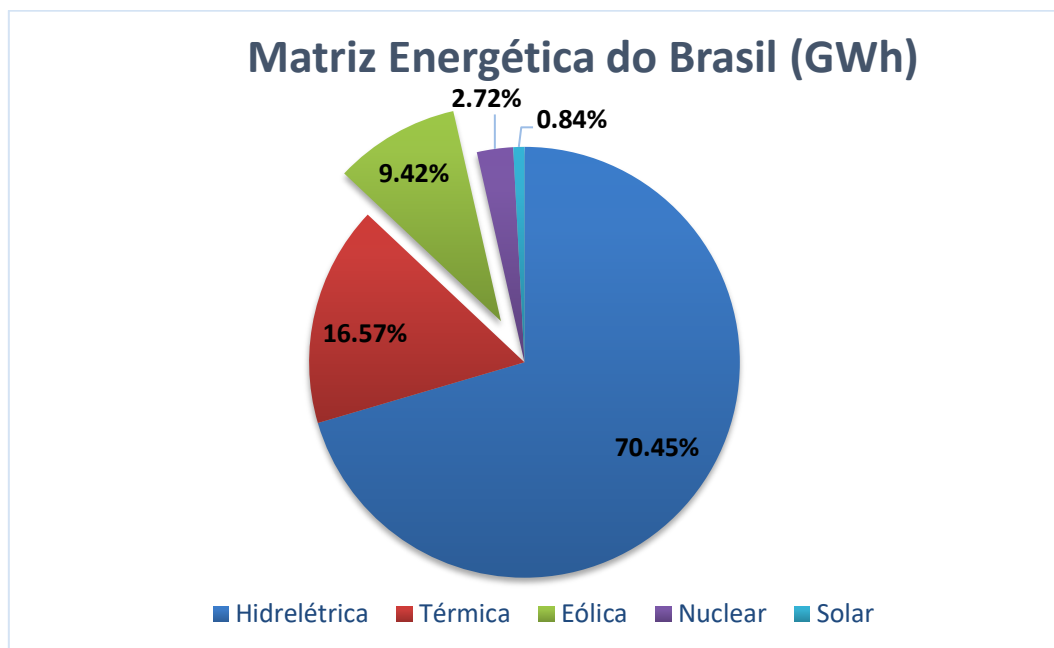


Figura 3 - Matriz Elétrica Nacional.

Fonte: Elaborado a partir de dados do ONS (2020)

No ano de 2019, 9,42% da geração de energia elétrica do Brasil foi proveniente de fonte eólica, o que representa um percentual relativamente alto comparado com a segunda fonte mais utilizada, a térmica, que teve 16,57% do total (ONS, 2020).

Nesse contexto, a energia eólica vem se inserindo como uma fonte complementar de energia no Brasil, e seu crescimento está também relacionado como uma forma de auxiliar as fontes térmicas, e diminuir o seu uso nos períodos de escassez hídrica, o que aos poucos já está acontecendo, como pode ser observado na Figura 4, onde é possível visualizar que no decorrer dos anos os períodos onde houve pouca produção por meio da fonte hidrológica houve um aumento na produção vindo de fonte eólica, apesar deste ainda não ser suficiente para suprir a demanda proveniente da falta das chuvas nos períodos de seca, mas já se apresenta como uma fonte complementar, que no futuro tem grandes chances de contribuir de forma significativa com a demanda necessária visto o seu rápido crescimento nos últimos anos (Figura 5).

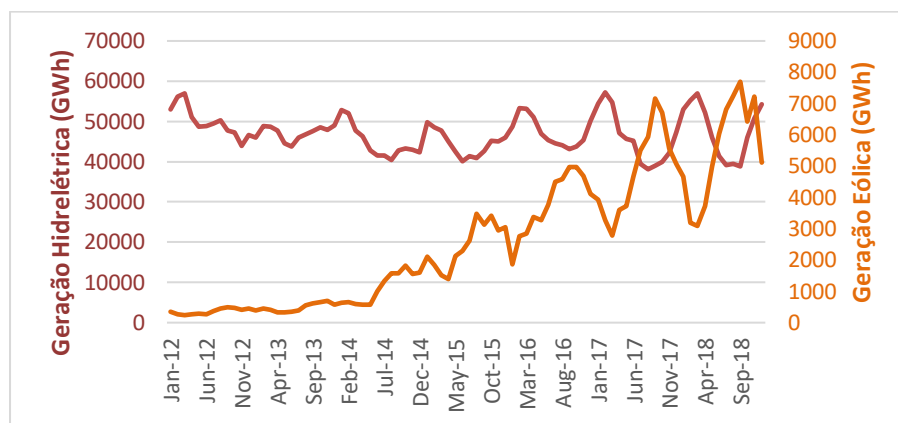


Figura 4 - Relação entre geração hidrelétrica e geração eólica

Fonte: Elaborado a partir de dados do ONS (2020)

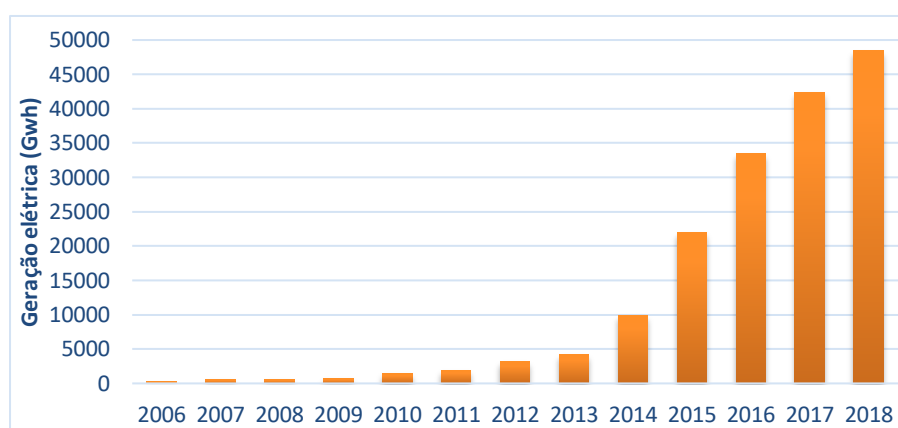


Figura 5 - Geração de Energia Eólica (GWh)

Fonte: Elaborado a partir de dados do ONS (2020)

A velocidade do vento é intermitente na natureza e, assim, sua geração tem comportamento estocástico (Aasim *et al.*, 2019). O que tem sido um grande problema quanto a maior inserção da fonte eólica e o seu crescimento é o fato de que ao contrário das fontes hídricas que contam com reservatórios para estoque, o vento não pode ser estocado. Então, a sua penetração na rede elétrica traz vários desafios para a estabilidade do sistema. Uma inesperada variação na produção de energia eólica pode aumentar a incerteza para o sistema elétrico, o que requer uma previsão de alta precisão das propriedades desse tipo de energia (Eissa *et al.*, 2017).

Nesse contexto, se torna muito difícil prever com precisão a energia eólica por causa de suas características estocásticas (Liu *et al.*, 2018). Essa intermitência do vento pode acontecer dentro de um mês, de uma semana ou até mesmo dentro de um dia, como pode ser visualizado na Figura 6, ao ser coletado

dados da velocidade do vento no decorrer de 24h de três dias diferentes, em que, para um mesmo horário, um dia pode apresentar uma velocidade maior, e no outro pode quase não existir.

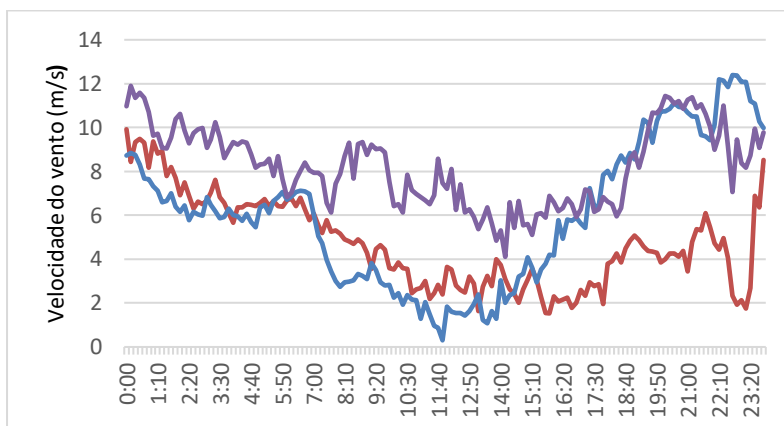


Figura 6 - Velocidade do vento no decorrer do dia (em m/s)
Fonte: Empresa do setor elétrico (2020)

Sendo assim, apesar da potência resultante da velocidade do vento ser um dos recursos mais populares de baixo carbono, e contribuir não apenas para a conservação de energia, mas também para a proteção ambiental, a previsão da velocidade do vento se torna um problema crítico na geração de energia eólica (Zhao *et al.*, 2016), e deve ser analisado de forma mais precisa para que sua implementação seja feita de forma mais completa no sistema elétrico de um país.

A previsão de energia eólica a curto prazo é um campo de pesquisa extremamente importante para o setor de energia, pois os operadores do sistema devem lidar com uma quantidade importante de energia flutuante devido ao aumento da capacidade instalada de energia eólica (Landberg, 2001) e com a crescente integração da energia eólica na rede elétrica, surgem novos desafios devido à sua intermitência e volatilidade. Deste modo, boas ferramentas de previsão desempenham um papel fundamental para a solução desses desafios. Em muitas áreas da geração eólica, as previsões do vento com um a dois dias de antecedência se tornaram importantes, incluindo a energia eólica, onde as empresas elétricas com alta penetração dessa energia precisam dessas informações em sua programação (Landberg, 2001). Além disso, a previsão de vento de curto prazo também é importante para garantir a estabilidade de todo o sistema elétrico nos casos de grande penetração da energia eólica na rede (Troncoso *et al.*, 2015).

Velocidade e geração eólica estão intimamente ligadas, então como a energia eólica disponível a qualquer momento depende da velocidade do vento

naquele momento, que é uma variável aleatória (Patel, 2005), a precisão da previsão da velocidade do vento influencia diretamente a previsão da geração eólica.

Diante deste contexto, visando contribuir para a melhoria da previsão da velocidade do vento para geração de energia eólica no período de curto prazo, este trabalho tem como objetivo fazer a previsão de velocidade do vento aplicando o método de previsão *Singular Spectrum Analysis* (SSA) a uma usina eólica localizada no Brasil, e a partir disso fazer uma análise comparativa de modelos SSA considerando diferentes horizontes de previsão e conjunto de treinamento, para diferentes dias de previsão, com diferentes tamanhos de série temporal utilizando as métricas RMSE e SMAPE. Para isso, é realizada uma comparação da série temporal de um ano todo, com somente o último mês desta série, prevendo os últimos sete dias do mês de dezembro, não sendo utilizada a comparação com diferentes métodos de previsão, e sim uma comparação com diferentes modelos SSA.

O método *Singular Spectrum Analysis* é usado para reduzir o ruído e extrair as informações de tendência dos dados originais de velocidade do vento (Mi *et al.*, 2019). Também funciona bem para séries temporais lineares e não lineares, estacionárias e não estacionárias, com diferentes características e estrutura. Ele pode identificar e extrair eficientemente os componentes de tendência e ruído de uma série temporal e depois reconstruir uma nova série, eliminando os componentes de ruído e, portanto, melhorando o desempenho da previsão. Com essa característica, o SSA é normalmente empregado para filtragem de série temporal no estágio de pré-processamento (Gao *et al.*, 2016), deste modo, o SSA, como um novo método analítico, é especialmente adequado para a pesquisa de oscilações periódicas, que provou ser uma ferramenta eficaz para análise de séries temporais em diversas aplicações, e os resultados indicam que ele pode efetivamente remover o ruído dos dados de velocidade do vento para melhorar o desempenho da previsão (Du *et al.*, 2016). A utilização da metodologia SSA foi escolhida por ser um método ainda pouco abordado e inovador, que obtém resultados precisos e já utilizado em outros cenários obtendo resultados promissores.

O estudo encontra-se organizado da seguinte forma: neste capítulo, o tema da dissertação foi abordado em uma introdução, seguido da justificativa do mesmo, e também dos objetivos da pesquisa; o capítulo dois, consiste de uma concisa revisão de literatura acerca do tema abordado, trazendo o que há de mais novo na literatura sobre o método utilizado; o terceiro capítulo, é composto pela

apresentação do método SSA, a descrição das suas etapas, e embasamento teórico; o quarto capítulo faz a apresentação dos dados que são utilizados no estudo fazendo uma breve análise descritiva da base de dados, seguido dos resultados obtidos com a aplicação do método nas duas situações e posteriormente análises comparativas. Por fim, o quinto capítulo apresenta as conclusões que o estudo proporcionou além de sugestão para pesquisas futuras, seguido das referências bibliográficas utilizadas para a realização do mesmo.

2 Referencial Teórico

Neste capítulo são abordados os principais conceitos referentes aos temas que embasaram a construção dessa dissertação. A seção 2.1 é referente a previsão de séries temporais; a seguir, a seção 2.2, apresenta trabalhos na área de previsões realizados no horizonte de curto prazo, e por fim, a seção 2.3, aborda as principais aplicações da técnica SSA nos diversos campos incluindo o setor de energia eólica.

2.1 Previsão de Séries Temporais

Existem diversos métodos de previsão de séries temporais, estes podem ser baseados desde simples intuição até modelos quantitativos e complexos. Nesta pesquisa, os modelos são utilizados em uma base de dados que contém as características passadas observadas e servem de guia para fazer afirmações além do período de conhecimento do fenômeno observado.

Em relação a classificação quanto aos horizontes temporais, Liu *et al.*(2018) classifica a previsão do vento em quatro categorias:

- Previsão de curtíssimo prazo: de alguns segundos a quatro horas à frente;
- Previsão de curto prazo: até três dias à frente;
- Previsão de médio prazo: a partir de três dias a mês à frente;
- Previsão de longo prazo: um ano à frente.

Dependendo do tempo e da escala de interesse, diversas técnicas podem ser utilizadas para a previsão da velocidade do vento e estão dispostas na literatura. Algumas podem fornecer uma previsão mais precisa para determinada escala de tempo específica, enquanto podem ser mais eficazes para outra, ou aplicáveis com maiores performances para outros objetivos de estudo, e interesses da atividade em questão.

Os métodos disponíveis na literatura são, no geral, categorizados como físicos e estatísticos (Filik, 2016). No entanto, estes métodos podem se subdividir formando 4 categorias principais apresentadas por Soman *et al.* (2010) e classificados na maioria dos estudos como: persistência, físico, estatístico e híbrido. Entretanto, durante a revisão da literatura, identificou-se mais uma categoria que pode ser classificada como métodos de previsão de velocidade de vento, os modelos espaço-temporal. Deste modo, este trabalho irá considerar 5 categorias nesta revisão:

- Método de persistência: Método também conhecido como '*Naive Predictor*'. Nele é assumido que a velocidade do vento futura é idêntica ao último valor obtido.
- Métodos com abordagem física: Métodos com abordagem física consistem em vários sub-métodos que, juntos, fornecem a transcrição da previsão do vento em um determinado ponto e nível de modelo, para prever a potência em um local (Nandana Jyothi & Ramana Rao, 2018). Métodos físicos usam parametrizações baseadas em uma descrição física detalhada da atmosfera, dentre eles estão os modelos de Previsão Numérica do Tempo (do inglês *Numeric Weather Prediction* - NWP), que são os mais encontrados na literatura quando se trata de métodos físicos, operam resolvendo modelos matemáticos complexos que usam dados climáticos como temperatura, pressão e rugosidade da superfície (Soman *et al.*, 2010);
- Métodos estatísticos: Estes métodos não são baseados em nenhum modelo matemático predefinido e, sim baseados em padrões. Os erros são minimizados se os padrões forem atendidos de acordo com os históricos. De acordo com Soman *et al.* (2010) uma subclassificação desta abordagem pode ser feita: modelos baseados em séries temporais e métodos baseados em Redes Neurais (RN);
- Métodos híbridos: Em geral, a combinação de diferentes abordagens, como a mistura de abordagens físicas e estatísticas ou a combinação de modelos de curto e médio prazo, é considerada uma abordagem híbrida (Soman *et al.*, 2010).
- Métodos espaço-temporais: Recentemente, os modelos de correlação espacial, também conhecidos como métodos "espaço-temporais", aparecem como uma nova tendência na previsão de velocidade do vento a curto prazo (Filik, 2016). Estes modelos já podem ser vistos em diversos trabalhos na literatura relacionados a previsão de energia eólica de

curtíssimo prazo, como em Cavalcante *et al.* (2017), Zhao *et al.* (2018), Dowell & Pinson (2016). Esses métodos consideram as séries temporais da velocidade do vento dos locais previstos e seus locais vizinhos para prever a velocidade do vento, esses modelos apresentam alguns desafios, como a necessidade de medir e transmitir a velocidade do vento aos locais correlacionados.

2.2

Previsão da velocidade do vento de curto prazo

O estudo está voltado para o problema de operação no setor elétrico para curto prazo, visto que os dados disponíveis estão em base horária, além da importância desse horizonte de tempo influenciar na operação das primeiras horas, ocasionando maior planejamento para as concessionárias de energia. Para este intervalo de tempo, alguns métodos são mais utilizados.

Após uma busca inicial, foram identificados estudos aplicando diversos métodos de previsão em base de dados para o horizonte de curto prazo, com isso constatou-se que se tratando de previsão de curto prazo a literatura aborda os mais diferentes métodos nas mais diversas aplicações.

Os modelos ARMA e ARIMA são os modelos mais populares usados para prever a velocidade do vento (Catalão *et al.*, 2011). A previsão de curto prazo da velocidade do vento pode ser baseada principalmente em dados históricos, através de alguns modelos lineares, como o modelo de média móvel autoregressiva (Erdem & Shi, 2011), neste estudo é realizada a previsão da velocidade do vento e da tupla de direção através de quatro abordagens baseadas no Método da Média Móvel Autoregressiva (ARMA).

Cadenas *et al.* (2016) comparou dois modelos de previsão de velocidade do vento, um deles foi um modelo univariado desenvolvido usando uma Média Móvel Integrada Autoregressiva Linear (ARIMA), o outro um modelo multivariado desenvolvido usando uma rede neural artificial exógena autoregressiva não linear (NARX); já Cadenas *et al.* (2010) fez a previsão do último dia das medições utilizando o método de suavização exponencial. Os resultados mostraram que o método de suavização exponencial é uma boa alternativa para a previsão de vento para os dados analisados. Segundo os autores, para previsões de velocidade do vento a curto prazo, esta técnica responde de maneira satisfatória às necessidades de precisão, devido à sua simplicidade e exatidão em comparação

com outras técnicas como Redes Neurais Artificiais (RNA), essa técnica mostrou-se muito útil para a previsão da velocidade do vento.

No entanto, outras abordagens mais complexas são feitas, por exemplo, no campo da inteligência artificial como foi trabalhado em Troncoso *et al.* (2015) que utilizou diversos tipos de árvores de regressão para previsão da velocidade do vento buscando avaliar o desempenho desses diversos modelos, incluindo algoritmos lineares e não-lineares; já Guo *et al.* (2012) propôs um modelo modificado de Redes Neurais *Feed-Forward* (FNN) para previsão de velocidade do vento, modelado através da melhoria da rede neural artificial baseada em EMD, a fim de remediar as deficiências, para a previsão da velocidade do vento; Cadenas & Rivera (2009) aplicou a técnica de Redes Neurais Artificiais (RNA) para prever a velocidade do vento à uma série horária, neste estudo, diferentes modelos de RNA foram desenvolvidos para a previsão de velocidade do vento em curto prazo no México, o que obteve bons resultados comparados aos métodos testados.

Tian *et al.* (2019) apresentou uma máquina de aprendizado extremo minimizada por erro (EM-ELM) especialmente adequada para a previsão de velocidade do vento a curto prazo devido à sua boa capacidade de mapeamento não linear combinado com algoritmo de otimização de enxame de partículas (PSO) para otimizar o algoritmo ELM. Portanto, o EM-ELM otimizado pelo algoritmo PSO aprimorado apresenta melhor desempenho de previsão.

Gani *et al.* (2015) apresenta uma nova abordagem híbrida, integrando a máquina de vetores de suporte (SVM) com o algoritmo de vaga-lume (FFA), Sun *et al.* (2015) propõe um modelo de decomposição de modelo empírico de conjunto rápido modificado (FEEMD) juntamente com um algoritmo de batente (BA) associado a máquinas de vetores de suporte à natureza (LSSVM) chamado de (FEEMD-BA-LSSVM) combinado com entrada selecionada por dados quantitativos profundos de análise. As séries originais de velocidade do vento são primeiro decompostas em um número limitado de funções do modo intrínseco (FMI) com uma série residual. Estes estudos combinam métodos de forma a obter previsões cada vez mais precisas.

Utilizando o método espaço-temporal, Cavalcante *et al.* (2016) realiza um estudo utilizando esse método através de técnica de previsão escalável baseado em computação paralela, algoritmo de otimização de convergência rápida e cálculos matriciais, através de dados horários para previsão de curto prazo. Os resultados encontrados foram comparados a abordagem de um processo autorregressivo vetorial esparso (sVAR) em termos de vantagens e limitações;

Filik (2016) faz uso de modelos lineares espaço-temporais (multicanais), que usam as medidas vizinhas ao redor da localização para investigar um problema de previsão de velocidade do vento a curto prazo. A formulação do problema dos modelos ARMA multicanal (denominada MARMA) é apresentada e técnicas eficientes de estimativa de coeficientes de previsão multicanal são revisadas.

Dentre os diversos métodos citados, observa-se a presença de métodos híbridos combinando duas ou mais abordagens em busca de uma previsão mais precisa, principalmente em trabalhos mais atuais. Uma vertente desses modelos híbridos é a combinação entre métodos de previsão com métodos de decomposição para curto prazo buscando melhores dados para previsão, como pode-se notar em De Giorgi *et al.* (2014) que aborda um método híbrido combinando Máquina de Vetor de Suporte de Mínimos Quadrados (LS-SVM) com Decomposição *Wavelet* (WD) e compara com uma abordagem baseada com métodos híbridos baseados em Redes Neurais Artificiais (RNA) para a previsão da energia eólica produzida por um parque eólico localizado no sul da Itália para vários horizontes de tempo, de uma hora a um dia. Os resultados do método híbrido superam os demais comparados.

Catalão *et al.* (2011) faz uma combinação utilizando a transformada *wavelet* combinada a redes neurais para previsão de energia eólica a curto prazo em Portugal. O método apresentou bons resultados comparados às abordagens persistência, ARIMA e de Redes Neurais.

Por fim, mais referências podem ser encontradas nos trabalhos de Qian *et al.* (2019) que faz uma revisão sobre métodos de decomposição com métodos híbridos para previsão de energia eólica; Lei *et al.* (2009) faz uma revisão dos principais métodos utilizados para a previsão da velocidade do vento para curto prazo, com uma abordagem ampla que abrange quais as características de cada um, vantagens e desvantagens; Zhu & Genton (2012) fazem uma revisão sobre alguns modelos clássicos de previsão estatística de séries temporais, incluindo filtros *Kalman* e modelos ARMA, além de abordarem alguns modelos espaço-temporais; Zhang *et al.* (2014) faz uma revisão detalhada dos métodos de ponta e novos desenvolvimentos na previsão de incerteza em energia eólica.

Diante das análises feitas, um dos *gaps* encontrados é a pequena quantidade de aplicações da técnica SSA principalmente como método de previsão, especialmente quando relacionado a estudos eólicos. Dentre os estudos realizados com a técnica é de destaque os trabalhos de Vargas (2015) e Cassiano (2014), Aguilar Vargas (2015) utiliza a técnica SSA para previsão validando-a em uma série temporal das médias horárias da velocidade do vento e da produção

eólica de um parque eólico no Brasil, já Cassiano (2014) propõe o uso de um método de Clusterização Baseada em Densidade, o DBSCAN (*Density Based Spatial Clustering of Applications with Noise*) na terceira fase do método SSA de modo a separar os componentes de ruído na fase de agrupamento das autotriplas do SSA nas séries temporais, buscando melhorar a previsão com esta técnica.

Deste modo, torna-se um campo de interesse a aplicação da técnica buscando a comparação dos resultados com este método no campo da previsão da velocidade do vento para curto prazo de forma a agregar a literatura deste meio. Outro *gap* encontrado é a ausência de estudos realizados no Brasil para o horizonte de curto prazo, sendo encontrados poucos trabalhos relacionados ao país, como o de Aguilar Vargas (2015).

2.3

Singular Spectrum Analysis (SSA)

A análise de espectro singular, do inglês *Singular Spectrum Analysis* (SSA), é um método para análise de sinais e é uma alternativa não-paramétrica aos métodos clássicos de séries temporais. Essa metodologia provou ser eficaz na análise de séries não estacionárias e complexas, uma vez que é um método não paramétrico e não requer as premissas clássicas sobre a estacionariedade ou sobre a normalidade dos resíduos.

Muitos estudos já foram realizados utilizando a técnica SSA, nos mais diferentes campos. No entanto, no campo de energia, pouco se vê sobre o uso da técnica, ainda mais quando se aborda sobre energia eólica. Os campos em destaque com uso da técnica são, economia e finanças, na previsão de variáveis financeiras e econômicas, aplicações no setor industrial (Hassani *et al.*, 2009); estudos do PIB (Hassani *et al.*, 2010; De Carvalho *et al.*, 2012; Rodrigues & Mahmoudvand, 2018; Hassani *et al.*, 2019); previsão de inflação (Hassani *et al.*, 2012). No campo do turismo algumas aplicações também são observadas (Beneki *et al.*, 2012; Hassani *et al.*, 2017) (Beneki *et al.*, 2012; Hassani *et al.*, 2017); abordagens em energia como em Beneki & Silva (2013) e Silva (2013); além de aplicações em geofísica (Majumder & Kanjilal, 2019), e mais recentemente aplicações no ramo do comportamento do consumidor em Silva *et al.* (2019) podem ser encontradas utilizando a técnica em questão.

Algumas variações do método são encontradas na literatura como o MSSA (Hassani *et al.*, 2012; Rodrigues & Mahmoudvand, 2018), e o SSA randomizado, uma variação do método original, alternativo para aplicações em séries longas

sem perder a qualidade das análises que é discutido por Rodrigues *et al.* (2018) e é comparado em termos de qualidade do ajuste e previsão do modelo e em termos de tempo computacional. Em um trabalho recente, Hassani *et al.* (2019) fez uma abordagem buscando resolver o problema de interpolação/ previsão de frequência mista utilizando o contexto do SSA.

Deste modo, é notável a vasta aplicabilidade da técnica nos mais diferentes setores relacionados tanto a previsão dos mais diversos dados como a filtragem dos mesmos se necessário, dependendo da situação. Sendo assim, uma aplicação no contexto de energia eólica seria pertinente, para avaliar a aplicabilidade do método e eficácia do mesmo diante dos métodos clássicos já abordados na literatura.

Quando se fala de SSA no contexto eólico, as aplicações estão relacionadas ao método de filtragem, raro são os trabalhos que o utilizam como um método de previsão. E, como foi visto nos estudos citados anteriormente, tal método apresentou bastante eficácia diante de diferentes métodos de previsão para diferentes horizontes.

As aplicações relacionadas a previsão da velocidade do vento, associam o método SSA a outros métodos de previsão, usando-o como um método de decomposição ou filtragem de dados melhorando os resultados previstos, como é visto em, Liu *et al.* (2018) utiliza o método de SSA antes de aplicar um método de *hashing* sensível a localização (LSH), para lidar com a alta volatilidade da série original, e para isso a decompõe em dois componentes: tendência média, e flutuação, onde revela as suas características estocásticas; Moreno & Santos Coelho (2018) apresentam uma aplicação relacionada ao Brasil, realizando a previsão de velocidade do vento usando o SSA, para que as saídas se tornassem mais estáveis, associado a um método de inteligência artificial conhecido como Sistema de Inferência *Fuzzy* Neuro Adaptativo obtendo melhores resultados no processo de previsão; Ma *et al.* (2017) aplica a análise de espectro singular para pré-processar os dados originais de velocidade do vento e assim obter uma sequência mais suave, e depois uma rede neural difusa dinâmica generalizada é utilizada para realizar a previsão.

Em relação às aplicações do SSA como método de previsão destaca-se: Wang *et al.* (2020) propôs um novo método híbrido de previsão baseado no método SSA e uma nova rede neural híbrida de Laguerre (HLNN), o SSA é usado para analisar as séries de energia eólica, que decompõem a série em duas subsequências, e uma nova rede neural híbrida de Laguerre é proposta para

construir o modelo híbrido de previsão de energia eólica otimizado pelo algoritmo de transição de estado de transição de oposição (OTSTA).

Yang & Wang (2016) apresentam no artigo um novo algoritmo que hibrida SSA, FAPSO (algoritmo *Firefly* e otimização de enxame de partículas) e RENN (rede neural rolante de Elman), para prever a velocidade do vento. O SSA é usado para decompor os conjuntos de dados de velocidade do vento originais em várias sub-séries. Então, a nova série é reconstruída. Os dados de velocidade do vento usados neste documento são tipicamente uma série temporal caótica, e o uso do SSA pode eliminar a influência de valores discrepantes e melhorar a precisão da previsão do modelo de previsão de velocidade do vento.

Yu *et al.* (2017) utiliza a WD para aumentar a precisão da previsão da velocidade do vento, para desagregar uma série original da velocidade do vento em várias sub-séries antes da previsão. No entanto, as sub-séries de frequência mais alta geralmente apresentam um grande distúrbio na previsão final. Deste modo é utilizada juntamente a técnica da SSA para fazer o processamento adicional nas sub-séries de frequência mais alta, em vez de fazer nenhuma modificação ou se livrar dela. Assim, é proposta uma tecnologia de decomposição híbrida chamada *Improved WT* (IWT). Depois disso é utilizado uma Rede Neural de Elman, criando um novo modelo híbrido IWT-ENN.

3

Singular Spectrum Analysis

A técnica *Singular Spectrum Analysis* (SSA) é uma poderosa técnica de análise de séries temporais que incorpora os elementos da análise clássica, estatística multivariada, geometria multivariada, sistemas dinâmicos e processamento de sinais (Hassani, 2007). A técnica possui como objetivo a decomposição da série de interesse em vários componentes aditivos que normalmente podem ser interpretados como componentes de tendência (isto é, suave e partes da série que variam lentamente), vários componentes oscilatórios (talvez com amplitudes variáveis) e componentes ruído (Golyandina *et al.*, 2001).

Ela é uma ferramenta muito útil que pode ser usada para resolver problemas como: encontrar tendências de diferentes resoluções; alisamento; extração de componentes de sazonalidade; extração simultânea de ciclos com períodos pequenos e grandes; extração de periodicidades com amplitudes variadas; extração simultânea de tendências e periodicidades complexas; encontrar estrutura em séries temporais curtas; e detecção de ponto de mudança (Hassani, 2007).

Esta técnica de decomposição é executada a partir de quatro principais etapas organizadas em dois estágios (Figura 7).

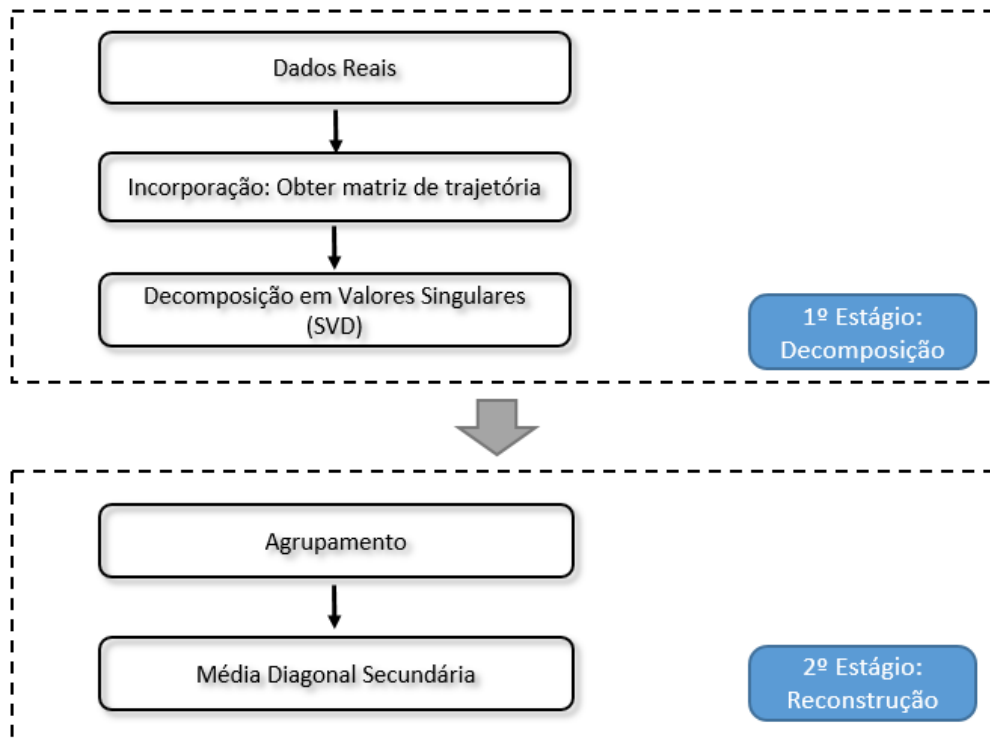


Figura 7 - Etapas da técnica SSA

Fonte: Elaborado com base em Golyandina *et al.* (2001)

Toda a metodologia foi baseada de acordo com Golyandina *et al.* (2001), as etapas e estágios da metodologia estão explicados em mais detalhes a seguir. Para maiores informações e outros exemplos, consultar Golyandina *et al.* (2001).

3.1

1º Estágio: Decomposição

Este estágio compreende as etapas de incorporação e decomposição em valor singular. Resumidamente, neste estágio a série temporal é decomposta em um conjunto de subséries. O único parâmetro necessário para realização desta etapa é a definição do tamanho de cada subsérie, também conhecido como tamanho da janela.

3.1.1

1ª Etapa: Incorporação

Seja $X = (X_0, \dots, X_{N-1})$ uma série temporal de tamanho N . A incorporação consiste no mapeamento da série temporal X em uma sequência de vetores defasados de tamanho L , tal que L é um inteiro e $1 \leq L \leq N$. L é conhecido como

tamanho da janela ou *window length* e representa a quantidade de componentes que a série X será decomposta. Segundo o estudo de Golyandina *et al.* (2001) o tamanho do L deve ser suficiente grande, mas não superior a $\frac{N}{2}$. O procedimento de incorporação forma $K = N - L + 1$ vetores defasados, com dimensão L .

$$X_i = (x_i, \dots, x_{i+L-1})^T \quad (1)$$

O resultado dessa etapa é a matriz de trajetória X , representada em (2):

$$X = \begin{pmatrix} x_0 & x_1 & x_2 & x_3 & \dots & x_{K-1} \\ x_1 & x_2 & x_3 & x_4 & \dots & x_K \\ x_2 & x_3 & x_4 & x_5 & \dots & x_{K+1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{L-1} & x_L & x_{L+1} & x_{L+2} & \dots & x_{N-1} \end{pmatrix} \quad (2)$$

Nota-se que a matriz de trajetória X apresenta uma estrutura semelhante a de uma matriz Hankel, que é simétrica e suas entradas $\{a_{ij}\}$ são constantes ao longo das diagonais paralelas à diagonal secundária. Os vetores defasados x_i são as colunas da matriz de trajetória X , as colunas e linhas de X são subséries da série original, deste modo podem ser comparadas a matriz de Hankel. Certamente, se N e L são fixos, há uma correspondência de um para um entre as matrizes de trajetória e a série temporal.

3.1.2

2ª etapa: Decomposição em Valor Singular (SVD)

A decomposição em valores singulares, ou *Singular Value Decomposition* (SVD) em inglês, é a segunda etapa do estágio de Decomposição. Nesse momento é feita a decomposição da matriz de trajetória X a partir da soma de matrizes elementares.

Seja $S = XX^T$, o produto entre a matriz trajetória e a sua transposta, e sendo denotados por $\lambda_1, \dots, \lambda_L$ os autovalores de S em ordem decrescente de magnitude ($\lambda_1 \geq \dots \geq \lambda_L \geq 0$), e por U_1, \dots, U_L os autovetores ortogonais e normalizados correspondentes da matriz de trajetória X . Desta forma, a decomposição em autovalores pode ser descrita como:

$$X = X_1 + X_2 + \dots + X_d \quad (3)$$

Sendo $X_i = \sqrt{\lambda_1} U_i V_i^T, i = 1, \dots, d$. Tal que $d = \max \{i, \text{tal que } \lambda_1 > 0\} = \text{posto}(X)$, isto é, d representa o número de autovalores diferentes de zero da matriz S ($d \leq L$). E, X_i representa uma matriz de posto unitário, que pode ser nomeada de matriz elementar.

A coleção ou tripla $\sqrt{\lambda_1}, U_i, V_i$ é chamada de autotripla (*eigentriple*) da matriz X , $\sqrt{\lambda_i}$ é chamado de valor singular da matriz X , e o seu conjunto representa o espectro de X .

3.2

2º Estágio: Reconstrução

O segundo estágio, chamado Reconstrução, tem como objetivo identificar grupos de matrizes mais semelhantes para a formação das componentes e posterior reconstrução da série. Este estágio é constituído pelas fases: agrupamento e média diagonal.

3.2.1

3ª etapa: Agrupamento

A primeira etapa do estágio de Reconstrução é chamada de Agrupamento (*Grouping*). Essa fase tem como objetivo dividir as matrizes elementares X_i em vários grupos e somar as matrizes dentro de cada grupo. O procedimento de agrupamento particiona o conjunto de índices $\{1, \dots, d\}$ em m subconjuntos distintos I_1, \dots, I_m . Seja $I = \{i_1, \dots, i_p\}$, então a matriz X_I correspondente ao grupo I , é definida como:

$$X_I = X_{i_1} + \dots + X_{i_p} \quad (4)$$

Onde $X_{i_1} + \dots + X_{i_m}$ são conhecidas como matrizes resultantes, após a decomposição, e estas são obtidas através da soma de matrizes elementares em um particular conjunto índice I_p .

O procedimento de escolher os conjuntos I_1, \dots, I_m é chamado de agrupamento das autotriplas (*eigentriple grouping*). Este fornece o último parâmetro da técnica SSA e único da etapa de Reconstrução, que é o número de autotriplas que serão agrupadas (r).

Em suma, na terceira etapa ocorre a divisão das matrizes elementares em diversos grupos e a soma dessas matrizes dentro de cada grupo. Deste modo,

tem-se o resultado desta etapa: a representação da matriz trajetória em forma de uma soma de matrizes resultantes.

3.2.2

4ª etapa: Média diagonal

Este é o último passo, conhecido como média diagonal (*diagonal averaging*), que transfere cada matriz I para uma série temporal, que é um componente aditivo da série inicial X .

Se X_{ij} significa um elemento de uma matriz X , então o k -ésimo termo da série resultante é obtido pela média x_{ij} sobre todo i, j tal que $i + j = k + 2$. Este procedimento é chamado de média diagonal, ou Hankelização da matriz X .

A transformação das matrizes resultantes em séries ocorre ao ser aplicado o operador linear de Hankelização (H). Esse operador age em uma matriz qualquer de modo a transformá-la em uma matriz de Hankel, e portanto, em uma matriz de trajetória, e por fim, em uma série. Em relação a esta etapa, este operador contabiliza as médias ao longo das linhas paralelas as diagonais secundárias das matrizes X_{ip} , para $p = 1, \dots, m$.

Seja T uma matriz $L \times k$ com elementos t_{ij} , $1 \leq i \leq L, 1 \leq j \leq k$. Sendo $N = L + K - 1, L^* = \min(L, k), k^* = \max(L, k)$, e $t_{ij}^* = t_{i,j}$. Ao fazer a média diagonal a matriz T é transformada em séries t_1, \dots, t_N usando a fórmula:

$$t_k = \begin{cases} \frac{1}{k} \sum_{m=1}^k t_{m,k-m+1}^* & \text{para } 1 \leq k < L^* \\ \frac{1}{L^*} \sum_{m=1}^{L^*} t_{m,k-m+1}^* & \text{para } L^* \leq k \leq k^* \\ \frac{1}{N-k+1} \sum_{m=k-k^*+1}^{N-k^*+1} t_{m,k-m+1}^* & \text{para } k^* < k \leq N \end{cases} \quad (5)$$

Isso equivale a fazer a média dos elementos da “antidiagonal” da matriz, $i + j = k + 1$. Então para $k = 1$ tem-se:

$$t_1 = t_{1,1}^*$$

Para $k = 2$

$$t_2 = \frac{t_{1,2}^* + t_{2,1}^*}{2}$$

E assim sucessivamente para $k = 3, 4, \dots, N$, até retornar ao formato original de série temporal.

3.3 Previsão com SSA

Pode-se obter predições através do método SSA em séries temporais que satisfazem (aproximadamente) às fórmulas recorrentes lineares (do inglês *linear recurrent formulae* – LRF). Deste modo, diz-se que uma série X_t satisfaz uma LRF de ordem $L - 1$ se os coeficientes a_1, \dots, a_{L-1} são tais que:

$$X_{N-j} = \sum_{p=1}^{L-1} a_p X_{N-j-p}; \quad \text{em que } 0 \leq j \leq N - L \quad (5)$$

Hassani (2007) afirma que os coeficientes podem ser encontrados através dos autovetores obtidos a partir do segundo passo, SVD. Outras informações em Golyandina *et al.* (2001), capítulo 2 e 5.

As séries regidas pelas LRF's podem ser bastante amplas, incluindo polinômios, exponenciais e termos harmônicos. Deste modo, uma série que satisfaz alguma LRF pode ser representada como a soma de produtos de polinômios, harmônicos e exponenciais, como a expressão (6).

$$X_t = \sum_{p=1}^q a_p(t) \varepsilon^{t\mu_p} \sin(2\pi t \omega_p + \varphi_p); \quad (6)$$

Onde $a_p(t)$ representam polinômios; μ_p , ω_p e φ_p são funções arbitrárias, e q representa o número de termos linearmente independentes.

Golyandina *et al.* (2001) apresentam a ideia de fórmula recorrentes lineares expressando séries governadas por uma LRF de ordem mínima, cuja dimensão é representada por d . De acordo com as características da decomposição SSA, existem no máximo d valores singulares não-nulos na SVD da matriz S ; então, mesmo que L e k sejam maiores que d , precisa-se de no máximo d matrizes elementares para reconstruir a série.

Danilov (1997) obteve o seguinte resultado: se o espaço linear E_r , conhecido também como espaço trajetória, gerado pelas colunas da matriz trajetória tem dimensão r menor do que L e este espaço é um espaço “não-vertical”, então a série satisfaz uma LRF com dimensão $L - 1$. Diante da suposição de espaço “não-vertical”, o espaço trajetória E_r , produz LRF's adequadas. Tais LRF's podem ser usadas para a previsão dos componentes da série, correspondendo às matrizes de posto unitário selecionadas na terceira etapa do método SSA. Esse processo de previsão é conhecido como algoritmo recorrente de previsão (do inglês, *recurrent forecasting algorithm*).

A previsão por SSA com $h = 1, \dots, M$ é obtida de acordo com a seguinte expressão recorrente:

$$\hat{X}_j = \sum_{p=1}^{L-1} a_p \hat{X}_{j-p}; j = N+1, \dots, N+M \quad (7)$$

Em que a_p são os coeficientes da combinação linear entre os $L-1$ últimos termos da série em questão. Deste modo, temos a seguinte expressão:

$$\hat{X}_{N+h} = \sum_{p=1}^{L-1} a_p \hat{X}_{(N+h)-p}; h = 1, \dots, M \quad (8)$$

Vale ressaltar que quanto maior o número de passos a frente (h), as previsões se tornam mais dependentes das previsões anteriores. Sendo assim, para que a predição tenha pelo menos um valor aproximado da série original, recomenda-se utilizar um horizonte de no máximo $M = L - 1$.

3.4 Informações Adicionais

A escolha dos parâmetros para a técnica SSA são feitos de acordo com o objetivo da análise e das informações iniciais da série temporal. Dentre elas, algumas informações adicionais auxiliam na escolha desses parâmetros para que sejam realizados de forma adequada de modo a obter o melhor resultado com a aplicação da técnica.

3.4.1 Tamanho de janela

Um parâmetro de suma importância é o tamanho da janela, e a literatura apresenta diversas informações que auxiliam na escolha do mesmo que é importante para a separabilidade dos componentes. Esse é o primeiro passo da técnica SSA, sendo assim o único parâmetro necessário para a etapa de decomposição, e um dos principais para a aplicação da técnica como um todo.

Deste modo, sua escolha é essencial para que seja realizado um bom agrupamento no passo seguinte, aumentando assim a precisão na aproximação com a série original.

Segundo Golyandina *et al.* (2001) valores entre $\left[2; \frac{N}{2}\right]$ para L são suficientes para realizar a decomposição da série temporal. No entanto, a decomposição pode se tornar mais detalhada na medida que esse valor é aumentado. Deste modo, nota-se que o comprimento ótimo de L (aproximação entre a série original

e a série reconstruída), coincide com o seu valor máximo estabelecido pelo intervalo, logo $L = \frac{N}{2}$.

No entanto, ao se trabalhar com séries que apresentem periodicidades em seus dados, é importante analisar tais sazonalidades de modo a escolher um comprimento de janela proporcional ao período de sazonalidade de forma a alcançar a separabilidade eficiente dos componentes (Golyandina *et al.*, 2001; Hassani, 2007).

Em todo caso, o L deve ser suficientemente grande para que cada vetor L – defasado obtenha o comportamento da série inicial. Para ilustrar essa situação Golyandina *et al.* (2001) apresenta um exemplo utilizando uma série de dados diários coletados por 14 anos, sendo assim, $N = 5113$ dados. Nesta série é possível observar dois comportamentos periódicos, sendo uma periodicidade semana ($T_1 = 7$) e uma periodicidade anual ($T_2 = 365$). Sendo T_2 muito maior que T_1 , é normal considerar L um múltiplo de T_2 abrangendo os dois comportamentos. Para o caso em questão, a escolha de $L = T_2$ garante a extração das duas periodicidades.

3.4.2

Escolha das autotriplas

O segundo parâmetro da técnica SSA que deve ser escolhido está relacionado a forma de agrupamento das autotriplas. Na terceira etapa, agrupamento, deve-se selecionar as autotriplas que vem da matriz X através da SVD de forma a constituírem grupos correspondentes a este passo.

No processo de formação dos grupos espera-se que haja uma boa separação entre os componentes da série temporal, sendo assim, envolve o problema da escolha das autotriplas que melhor se adequem entre todas as que são geradas pela série temporal. Deste modo, deve-se identificar as autotriplas correspondentes aos principais componentes da série temporal a ser reconstruída.

Independente do objetivo do estudo, as autotriplas são separadas em dois grupos, o sinal e o ruído, sendo o sinal constituído por todos os componentes de interesse e os demais componentes separados em um outro grupo, o ruído (Golyandina *et al.*, 2001).

O agrupamento das autotriplas depende também do propósito da análise, como mostrado por Golyandina *et al.* (2001), o efeito do agrupamento para diferentes objetivos como: extração da tendência, suavização da série,

agrupamento para obtenção dos componentes periódicos e agrupamento para obtenção de uma decomposição refinada da série.

Se o interesse é encontrar uma estrutura refinada da série, então as autotriplas correspondentes aos componentes de tendência e oscilações são tratados como sinal; e os componentes restantes são associados ao ruído. Ou seja, o ruído, assim como os resíduos da análise clássica de séries temporais, não deve ter uma estrutura de tendência e/ou periódica.

O comportamento dos vetores singulares, e dos valores singulares da SVD obtidos na matriz trajetória auxilia no agrupamento adequado das autotriplas. Deste modo é importante analisá-los também para resultados mais claros. A seguir são apresentados mais sobre essas ferramentas de análise.

3.4.3

Valores singulares

O comportamento dos valores singulares pode ser observado através de um gráfico que apresenta os valores de L em uma escala decrescente relacionada a magnitude ($\sqrt{\lambda_1} \geq \dots \sqrt{\lambda_L}$), expondo assim o espectro de X .

O primeiro valor singular, o que tem o maior valor absoluto sempre estará associado a componente de tendência. Do mesmo modo que a tendência em uma série temporal pode ser descrita como uma função com uma variação lenta ao longo do tempo, esse é o comportamento esperado do primeiro componente.

Em teoria, uma série que contenha somente resíduos produz uma sequência lenta decrescente de valores singulares. Desta forma é possível diferenciar as autotriplas relacionadas ao sinal, pois tem valores singulares altos, das autotriplas associadas ao ruído por seu comportamento, um exemplo desse tipo de gráfico está disposto na Figura 8.

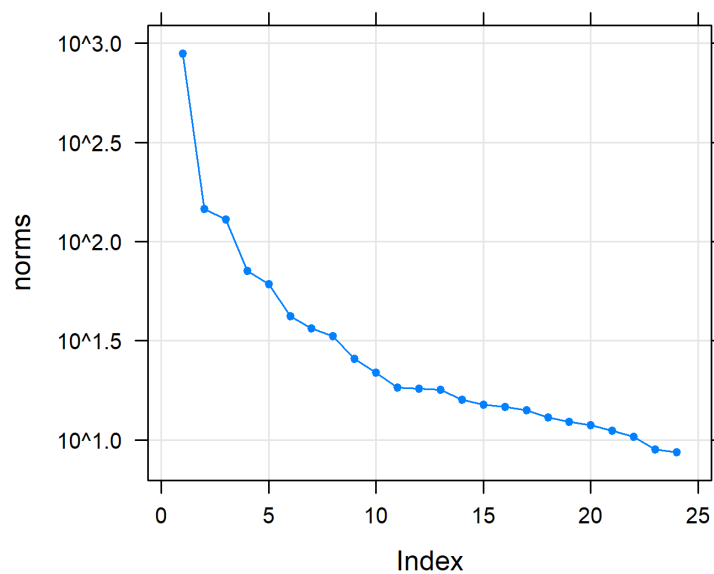


Figura 8 - Exemplo de gráfico de valores singulares

Fonte: Autora (2020)

Outra característica importante a ser analisada nesse tipo de gráfico é quando dois valores singulares se encontram próximos em escala de magnitude, aparentemente com a mesma escala, pode-se dizer que eles formam um par de autovalores e representam componentes oscilatórios.

3.4.4 Correlação ponderada

Uma condição necessária para a separabilidade (aproximada) de duas séries é a existência de correlação ponderada nula (ou próxima de zero) dos componentes reconstruídos. Deste modo, os componentes reconstruídos pertencem ao mesmo grupo de autotriplas.

Uma ferramenta utilizada para esse processo de agrupamento é conhecida como “matriz de correlação – w ”. Essa ferramenta faz uma análise gráfica da matriz de correlações ponderadas, que indica as correlações entre as componentes da SVD através de uma escala de cores variando do branco ao preto, correspondendo aos valores absolutos das correlações, de 0 a 1, pode ser visualizada na Figura 9 um exemplo de matriz de correlações – w .

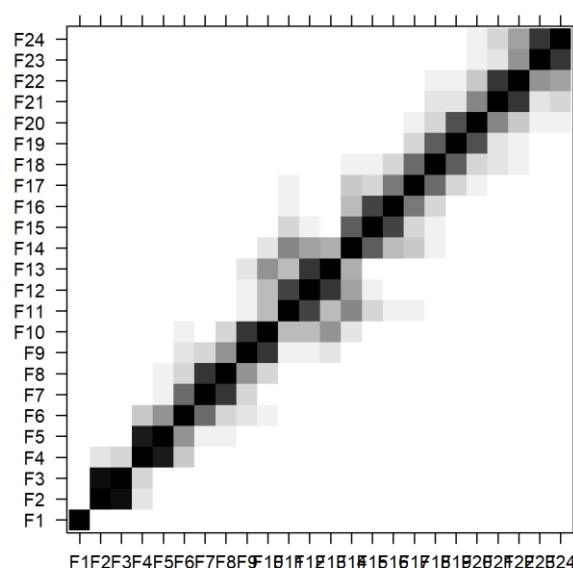


Figura 9 - Exemplo de matriz de correlações – w

Fonte: Autora (2020)

Um agrupamento adequado pode ser visto de acordo como é disposto na matriz de correlações w , onde espera-se visualizar uma divisão clara entre os componentes em duas partes, uma parte com tonalidades mais escuras de componentes mais correlacionados, indicando o grupo das autotriplas do sinal, e uma outra parte com tonalidades mais claras e mais dispersas, com componentes mais descorrelacionados, indicando o grupo das autotriplas do ruído. Na matriz utilizada como exemplo as componentes iniciais estão mais correlacionadas formando o grupo do sinal, e as demais formam o grupo ruído. Neste exemplo, nota-se a forte presença do ruído no decorrer da série, o que torna um pouco mais difícil essa visualização, sendo necessária o uso de outras ferramentas que agreguem informações para conclusões mais concisas sobre os parâmetros.

Após essa divisão, as componentes que formarão o sinal serão utilizadas para reconstruir a série original, que depois será prevista utilizando algum método de previsão.

4

Resultados

Neste capítulo, são apresentados os resultados obtidos a partir das aplicações do método de previsão proposto. Para isso, é feita uma análise descritiva da base de dados utilizada no estudo, duas aplicações seguidas de análises de sensibilidade do tamanho da janela, e por fim, uma comparação entre as aplicações realizadas para mostrar a eficácia e aplicabilidade do método estudado.

4.1

Descrição da base de dados

A base de dados utilizada para aplicação do método é oriunda de uma empresa do setor elétrico que contém 24 turbinas contendo dados de velocidade do vento dispostos no intervalo de 10 em 10 minutos. Para a aplicação do estudo foi utilizada somente uma turbina, e devido à altíssima frequência dos dados no intervalo, foi-se utilizada a média diária coletada para melhor aplicação da técnica. Deste modo, a base de dados utilizada no estudo dispõe de 8760 dados horários, no intervalo de 01 de janeiro de 2017 a 31 de dezembro de 2017. A série temporal com os valores das médias da velocidade diárias pode ser observada na Figura 10.

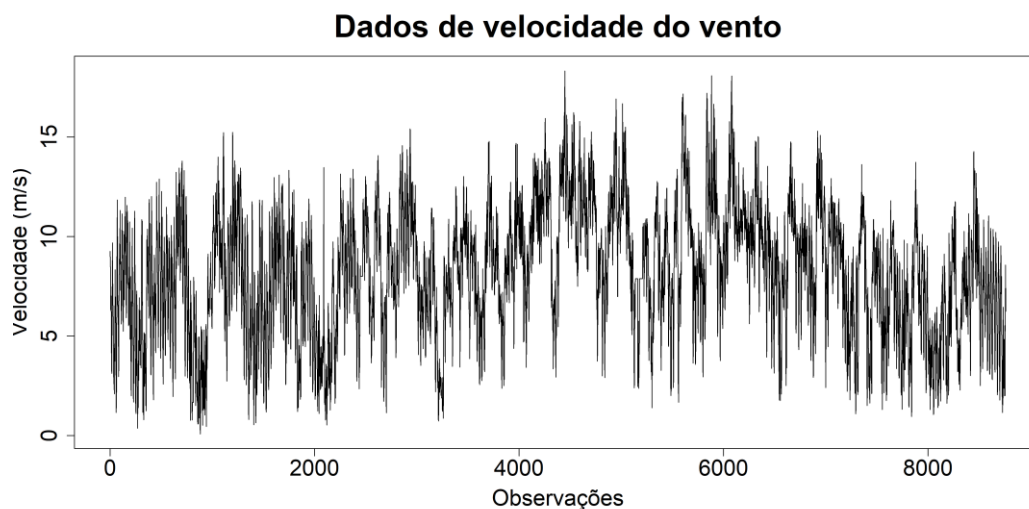


Figura 10 - Série temporal com as médias das velocidades diárias.

Fonte: Autora (2020)

A partir da Figura 10, nota-se um padrão com sazonalidades existentes característicos de séries temporais de velocidades de vento, além de, a partir da análise da série, percebe-se a existência de estacionariedade. Algumas informações estatísticas que caracterizam essa série temporal estão dispostas na Tabela 1, e podem ser visualizadas também na Figura 11, que é o *boxplot* da série mostrando a disposição dos dados.

Tabela 1 - Estatísticas da série temporal de velocidade do vento

Estatística	Valor
Tamanho	8760
Média	8,172
Mediana	8,329
Mínimo	0,075
Máximo	18,326
Desvio-Padrão	3,175
Assimetria	-0,084
Curtose	2,548

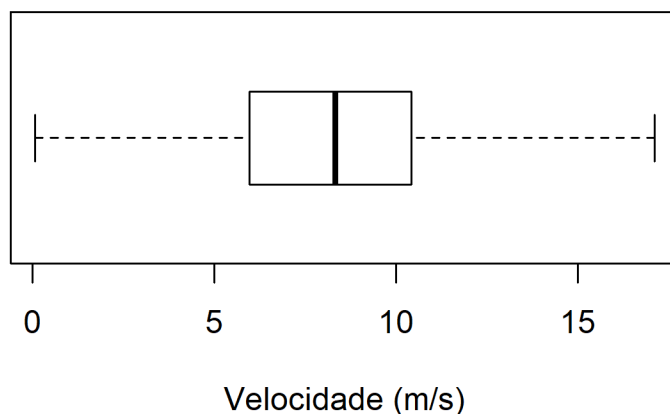


Figura 11 - *Boxplot* das observações

Fonte: Autora (2020)

De acordo com o *boxplot* apresentado na Figura 11, a maior concentração de dados de velocidade do vento aparece a partir de 6 m/s até 10 m/s, o que também pode ser visto na Tabela 1, com os valores da mediana próximos a 8. Com a análise do *boxplot* verifica-se uma grande variação nas velocidades, o que é visto na Tabela 1 que os valores podem estar entre 0 (mínimo) e aproximadamente 18 (máximo), o que já era esperado, visto que a velocidade do vento apresenta essa característica, onde determinadas horas do dia pode haver uma velocidade menor, e em outras uma velocidade maior, ou em determinados períodos do ano, o que pode ser observado na Figura 12 ao se analisar o *boxplot* com os dados de cada mês.

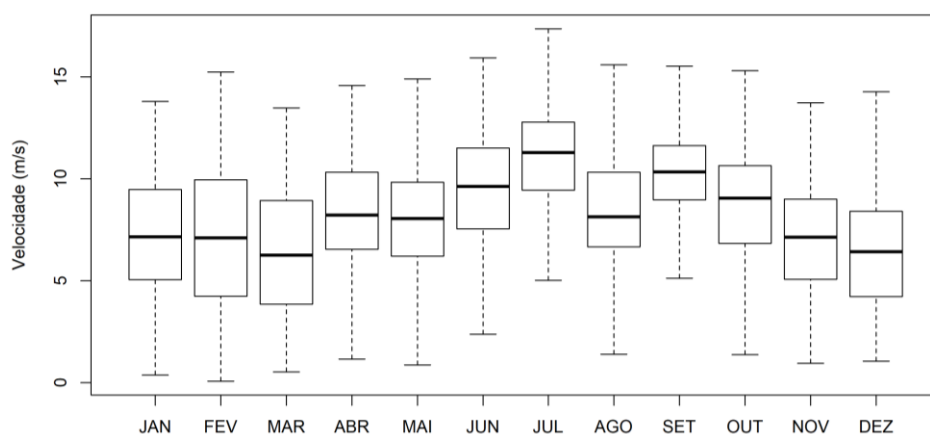


Figura 12 - *Boxplot* com os dados de cada mês

Fonte: Autora (2020)

Quando analisado o *boxplot* contendo as velocidades mês a mês (Figura 12), nota-se a oscilação entre as medianas de cada mês, apesar de os dados coletados pertencerem ao mesmo intervalo, o que confirma que as oscilações da

velocidade do vento compreendem valores entre 0 e aproximadamente 20, mas se concentram mais em valores entre 5 e 10, e em alguns casos como o mês de julho, podem ultrapassar esse valor, e meses como Março e Dezembro apresentarem valores menores que 5 em algumas situações.

A Figura 13 representa o histograma com os dados, reafirmando o que foi visto no *boxplot* e na tabela com as características.

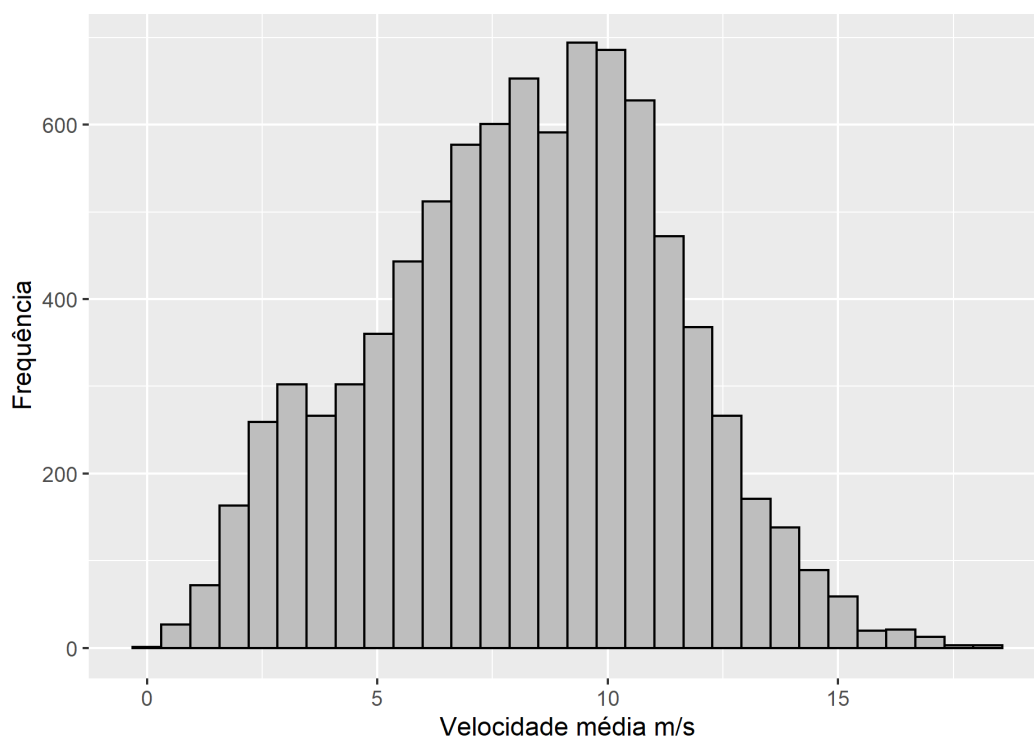


Figura 13 - Histograma com a frequência das velocidades

Fonte: Autora (2020)

A concentração dos dados encontra-se entre o intervalo de 7 e 12, onde pode ser visto apresentando frequências mais altas. Antes disso ou depois desse intervalo, a frequência começa a diminuir mostrando poucas ocorrências de velocidades mais baixas, assim como de velocidades mais altas, que podem surgir devido a situações atípicas ou outras características que podem interferir nos dados. Observa-se uma assimetria levemente negativa o que pode indicar que a distribuição dos dados analisados tem a “cauda” um pouco mais longa para a esquerda. Mas por esse valor ser bem próximo de zero, chega a ser similar a um padrão assimétrico, sendo quase imperceptível. Quanto a curtose, observa-se que o valor encontrado para o coeficiente é maior que o valor definido para o coeficiente da distribuição normal usada para comparação, deste modo, temos

que a distribuição de dados em questão tem uma medida de curtose maior, isto é, é leptocúrtica.

Após da análise descritiva da base de dados, a Figura 14 apresenta as etapas percorridas para a realização desse estudo e obtenção dos resultados a serem analisados.

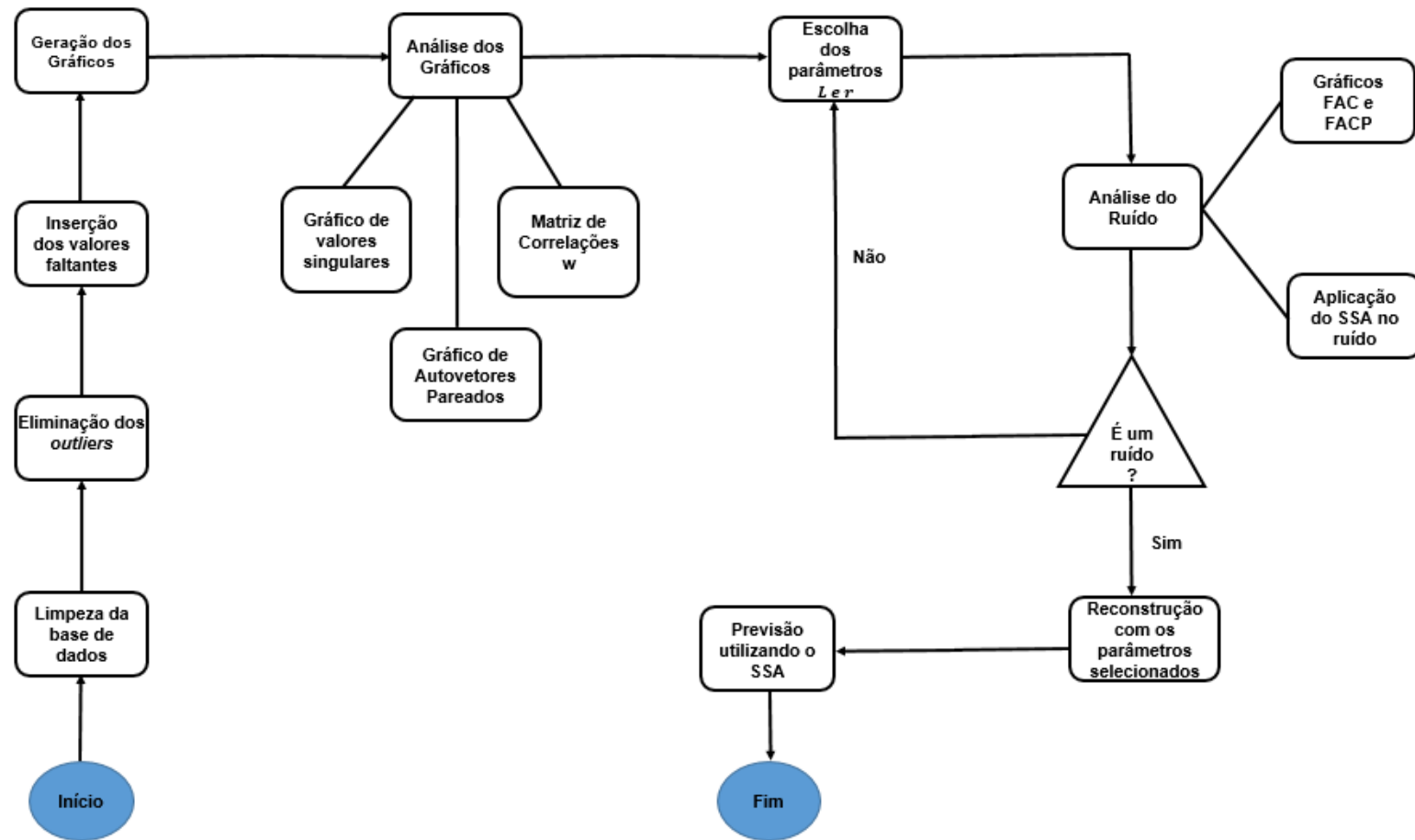


Figura 14 - Fluxograma com as etapas para realização da pesquisa

Fonte: Autora (2020)

Primeiramente, foi realizada uma limpeza na base de dados, seguida da detecção dos *outliers* (valores de velocidades muito altas, fora da normalidade, ocorrentes devido a erros de medição), que foram neutralizados utilizando a média do mês, visto que essas ocorrências correspondiam a um dia inteiro de observações, e assim foram incluídos os valores faltantes. Após esta etapa, obteve-se a base de dados para a aplicação da técnica SSA.

Em seguida foi realizada a geração dos gráficos de valores singulares, matriz de correlação e autovetores pareados, a partir desses gráficos foi possível identificar o melhor par de parâmetros. Os valores de L foram previamente escolhidos de acordo com a periodicidade da série, como os dados analisados são horários, os valores escolhidos para L foram múltiplos de 24 suficientemente grandes para abranger todos os componentes da série para a reconstrução. Após a escolha desses valores, os gráficos foram gerados buscando identificar os melhores valores para r , e também sendo analisados os valores de L escolhidos observando-se que uma boa separação entre os componentes estava sendo feita.

Para a escolha dos valores de r foi-se utilizado principalmente o gráfico dos autovetores pareados, sendo escolhido o valor onde houvesse uma separação entre o sinal e o ruído, identificado através da formação de figuras geométricas. Deste modo, no par onde era identificada a última figura formada foram considerados componentes do sinal até este valor, e a partir dele as demais componentes foram consideradas como componente do ruído. Para este estudo foi utilizado os pares de autovetores até este último r , agrupando as componentes que estavam anteriores a ele formando a componente sinal.

Com a escolha dos parâmetros, em seguida foi feita a análise do ruído, buscando verificar se havia ou não interferência do mesmo na nova série reconstruída, para isso foram utilizadas duas ferramentas de análise: os gráficos da FAC e FACP e a análise do gráfico de autovetores pareados a partir da aplicação do SSA no suposto “ruído”.

O melhor par de parâmetros foi escolhido a partir dos valores obtidos pelas métricas de erro utilizadas na pesquisa, sendo elas o SMAPE e o RMSE. Deste modo, os pares foram testados e aquele que obteve valores menores foi considerado o mais adequado para a série em questão. Tais métricas foram escolhidas por serem comumente usadas principalmente em estudos de previsão relacionados ao método SSA, sendo usadas para aferir a qualidade do ajuste ao modelo, apresentando análises similares, sendo o RMSE como uma medida análoga ao desvio padrão, e o SMAPE medindo em forma de percentagem.

A série então foi reconstruída utilizando os parâmetros propostos, e em seguida prevista, se não novos parâmetros devem ser encontrados de modo a obter uma série com pouca interferência de ruído.

4.2

Aplicações

Para o estudo foram realizadas duas aplicações do método de previsão SSA descrita em mais detalhes nas subseções a seguir:

1. Na primeira aplicação foi utilizado o ano todo como conjunto de treinamento para prever variados dias de dezembro;
2. Na segunda aplicação é utilizado como conjunto de treinamento somente o mês de dezembro para prever diferentes dias do mesmo mês.

Para os dois casos em questão foi feita uma comparação variando o conjunto de treinamento para fazer a previsão dos últimos dias do mês de dezembro de acordo com o conjunto de pares de parâmetros fixados (tamanho de janela e número de autotriplas) encontrados em cada caso. A partir disso, de acordo com as métricas RMSE e SMAPE, uma comparação pode ser feita de qual modelo apresenta melhor desempenho para prever os últimos dias de dezembro, utilizando o ano todo como conjunto de dados, ou utilizando somente o mês de dezembro para cada dia analisado.

4.2.1

Primeira aplicação

A primeira aplicação seguiu o esquema apresentado na Figura 15, onde foram utilizados diferentes dias como alvo de previsão, variando somente o conjunto de treinamento de acordo com o dia a ser previsto. Deste modo, seguiu-se o esquema da Figura 15:



Figura 15 - Conjuntos de treinamentos para os dias de previsão

Fonte: Autora (2020)

Para a aplicação do método SSA neste conjunto de dados primeiramente deve-se realizar a decomposição que é a primeira etapa, para isto, um valor para L deve ser encontrado e para r , sendo eles os melhores valores de modo a dividir a série em componentes que se aproximem a série original ao ser reconstruída. O objetivo do SSA nesse estudo é fazer a previsão, trabalhando apenas com o sinal, isto é, remove-se o ruído.

Sendo assim, várias combinações entre os parâmetros L e r foram feitas afim de encontrar o melhor par de valores para a série em questão, fazendo com que a série original sofra a menor influência do ruído. Para alcançar esse resultado algumas análises foram realizadas a partir de técnicas utilizadas para tal questão. Foram usadas as ferramentas matriz de correlações w , o gráfico de autovalores, e o gráfico de autovetores pareados. A partir da análise dessas ferramentas em conjunto, o melhor valor para L e r pode ser encontrado. Essa análise foi ratificada a partir da métrica RMSE (*Root Mean Square Error*) e da métrica SMAPE (*Symmetric mean absolute percentage error*), visto que o objetivo do trabalho é fazer uma comparação entre os melhores modelos SSA para previsão para os últimos dias de dezembro. A seleção dos tamanhos de janela L foi feita de forma a compreender a periodicidade que ocorre na série, sendo assim, foram

selecionados alguns valores suficientemente grandes para capturar todas as características e que ao mesmo tempo fossem múltiplos de 24, tendo em vista que séries de vento apresentam essa característica de ter uma periodicidade de 24, quando são analisados dados horários.

Depois das combinações realizadas, os valores encontrados para os melhores pares de L e r para a série em questão foram dispostos na Tabela 2.

Tabela 2 - Valores de L e r testados

L	r
120	56
240	101
360	132
384	129
576	196
720	360

A melhor combinação dos valores L e r encontrada foi 384 – 129 por apresentarem visualmente melhor divisão entre as componentes, retirando ao máximo a influência do ruído da série ao ser reconstruída. Esse resultado foi encontrado a partir da observação de todos os gráficos para cada um dos pares de parâmetros testados. A Figura 16 apresenta o gráfico de autovalores para o par selecionado, nessa imagem não fica muito visível visto a grande quantidade de componentes que foram decompostas, então a visualização fica comprometida. Após isso fica difícil a identificação devido a visualização do gráfico. Também é difícil identificar onde inicia o decaimento, que segundo a teoria marca as componentes pertencentes ao grupo do ruído. Sendo assim, análises complementares de outras ferramentas que permitam uma melhor visualização são necessárias para confirmar os componentes já analisados, e encontrar aqueles ainda não identificados.

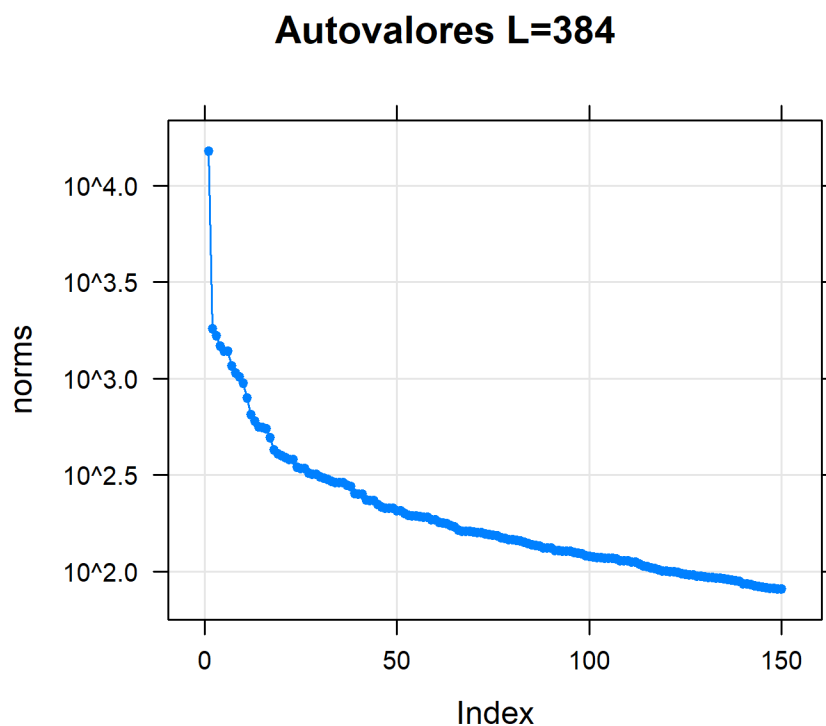


Figura 16 - Gráfico dos Autovalores

Fonte: Autora (2020)

Na matriz de correlações (Figura 17), a visualização também é um pouco comprometida devido ao grande número de divisões dos componentes, sendo assim, a análise dela sozinha é inconclusiva, necessitando do auxílio do gráfico de autovetores pareados para melhor identificação da melhor separação entre os componentes.

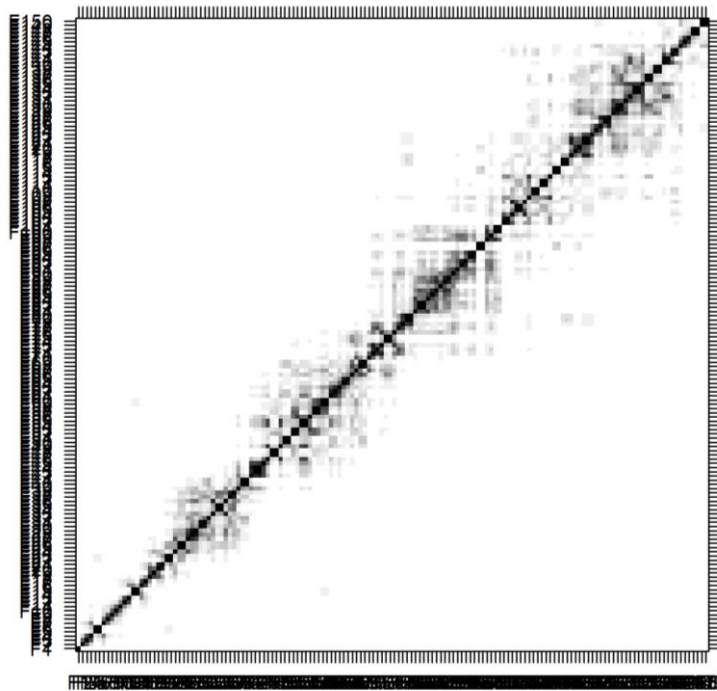


Figura 17 - Matriz de Correlações

Fonte: Autora (2020)

Para finalizar a análise, e trazer melhor visualização, foi observado o gráfico de autovetores pareados (Figura 18), que deu a visão completa das autotriplas que estão mais relacionadas ao sinal. Analisando para o comprimento de janela estabelecido, é possível observar que o valor de r relacionado a este comprimento L é 129, onde os autovetores pareados indicam ruídos, não fazendo menção a figuras geométricas posteriores a esta numeração. Também é possível observar que os componentes que estão bem relacionados e formam figuras geométricas bem definidas estão presentes nos pares de componentes já encontrados no gráfico de autovalores, confirmando a análise anterior.

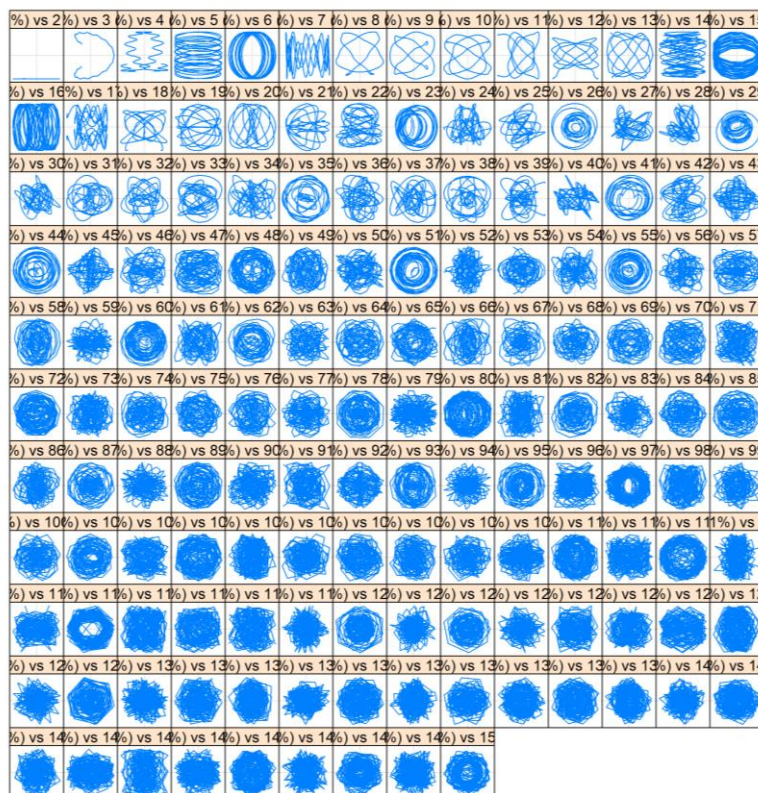


Figura 18 - Gráfico de autovetores pareados

Fonte: Autora (2020)

A partir dos parâmetros estabelecidos, a série foi reconstruída para que então pudesse ser prevista utilizando o algoritmo Rssa (*R Core Team*, 2018), que utiliza o método do Algoritmo Recorrente de Previsão, para fazer a previsão de 24 passos a frente, o que corresponde às 24 horas do dia seguinte.

A verificação da extração do ruído foi feita de dois modos: através dos gráficos das autocorrelações (FAC) e das autocorrelações parciais (FACP), e a partir da aplicação do método SSA no suposto “ruído”. Ambos resultados apontaram para a extração do ruído, deixando o mínimo de resíduo e correlação possível no sinal. A análise complementar no ruído com a técnica SSA foi feita através do gráfico de autovetores pareados, que comprovou um comportamento de componentes de ruído presentes na análise, ratificando a primeira análise.

O comportamento desenhado pelas previsões de acordo com os pares dos parâmetros selecionados pode ser observado na Figura 19, a qual trata-se do gráfico que compara cada previsão com o valor observado para o dia 31/12. Visualmente, pode-se concluir que todos os comprimentos de janela estabelecidos desenharam muito bem e acompanham os dados reais, deste modo realizando previsões próximas da realidade.

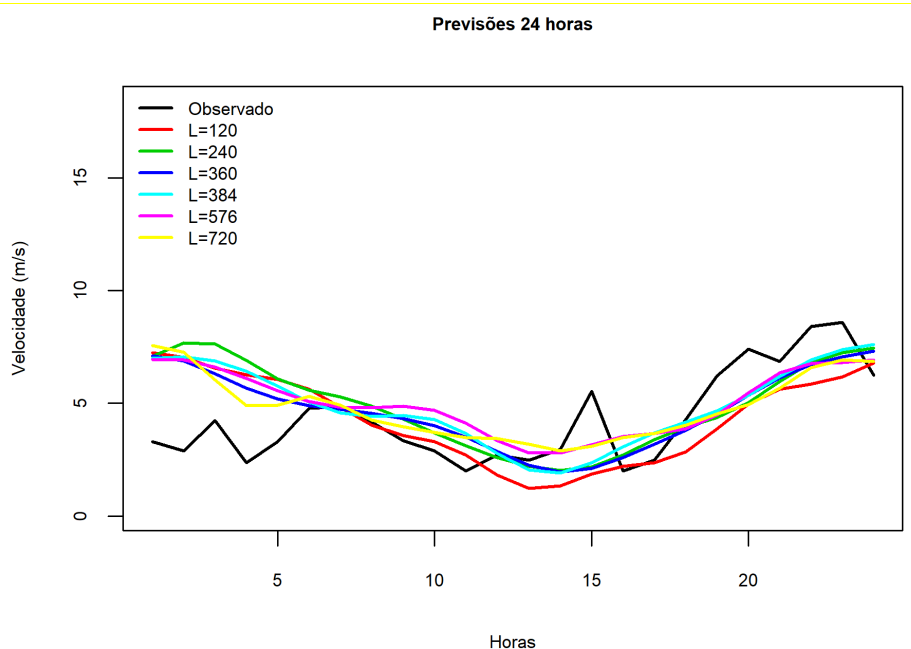


Figura 19 - Previsões de 24 horas para os tamanhos de janela para o dia 31/12

No entanto, é interessante também observar os valores encontrados de acordo com as métricas de erro para cada um dos pares analisados. A Tabela 3 apresenta a comparação entre as previsões feitas com as combinações dos valores L e r através das métricas RMSE e SMAPE. O menor valor obtido foi para $L = 384$ tanto para a métrica RMSE como para a métrica SMAPE.

Tabela 3 - Métricas para cada par de parâmetro testado

$L - r$	RMSE	SMAPE
120 – 56	2,168	39,06%
240 – 101	1,905	35,88%
360 – 132	1,715	32,68%
384 – 129	1,705	31,96%
576 – 196	1,912	34,59%
720 – 360	1,828	35,21%

Fonte: Autora (2020)

4.2.1.1

Análise de sensibilidade do tamanho L para a primeira aplicação

Os diversos tamanhos de janela testados para a primeira aplicação, com o conjunto de treinamento sendo todo o ano de 2017, e a previsão o último dia de

dezembro, se comportaram de forma a acompanhar os dados originais da série real, deste modo, observa-se que o conjunto com os $L's$ e $r's$ estabelecidos abrangem o conjunto de dados reais. Como forma de analisar o comportamento desses tamanhos de janelas, foi testado para outros dias de previsão os mesmos $L's$ e $r's$ já estabelecidos anteriormente, variando somente o conjunto de treinamento, para os últimos dias do mês de dezembro.

Sendo assim, através da Tabela 4, podemos ver, de acordo com os últimos dias de dezembro quais foram os melhores comprimentos de janela dentre os pré-estabelecidos.

Tabela 4 - Melhor par de parâmetros de acordo com o dia previsto

Treinamento	Previsão	$L - r$	RMSE	SMAPE
01/01 – 24/12	25/12	120 – 56	1,301	18,36%
01/01 – 25/12	26/12	120 – 56	1,026	14,73%
01/01 – 26/12	27/12	120 – 56	1,158	15,27%
01/01 – 27/12	28/12	576 – 196	1,366	22,03%
01/01 – 28/12	29/12	120 – 56	0,760	14,55%
01/01 – 29/12	30/12	240 – 101	2,653	49,30%
01/01 – 30/12	31/12	384 – 129	1,705	31,96%

De acordo com a métricas analisadas, os comprimentos de janela abrangeram de forma bem sucedida o intervalo em questão. As janelas 120 e 240 compreenderam a maioria dos dias, e nos dias que não obtiveram o melhor desempenho, também apresentaram valores próximos aos melhores obtidos no dia previsto. De modo geral, os pares selecionados adequaram-se ao intervalo previsto obtendo bons resultados.

4.2.2 Segunda aplicação

A segunda aplicação realizada para comparar com a primeira aplicação do método, foi feita na mesma base de dados, mas abrangendo somente o último mês observado, no caso dezembro. Deste modo, têm-se uma aplicação com menos dados, com um novo conjunto de treinamento que compreende o período de 01/12/2017 até 30/12/2017, totalizando 720 observações. Fazendo a previsão

em um primeiro momento do dia 31/12/2017, e em um segundo momento, fazendo um deslocamento dessa previsão para os dias anteriores, fazendo previsão de 24 horas, alterando somente o conjunto de treinamento utilizado os mesmos pares de parâmetros testados para a primeira situação, como pode ser visualizado na Figura 20. E após isso, a análise das métricas RMSE e SMAPE indicará qual o melhor par para o dia previsto. Após essas análises, uma análise comparativa geral pode ser feita, entre as duas aplicações.

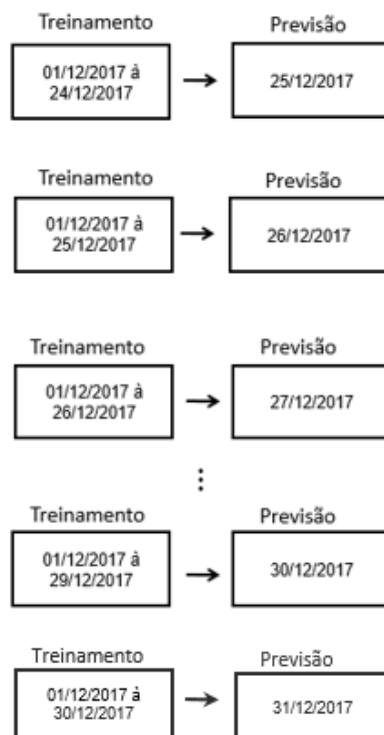


Figura 20 - Conjunto de treinamento para o dia previsto

Fonte: Autora (2020)

Como neste caso tem-se um novo conjunto de dados relativamente menor que no primeiro caso, foi necessário encontrar novos parâmetros L e r para serem ajustados a série de acordo com novas análises a serem feitas. Sendo assim, diante do novo intervalo, outras análises foram realizadas de modo a encontrar esses parâmetros que melhor se adequem a nova série temporal a ser decomposta e posteriormente reconstruída. Alguns pares foram propostos por critério de comparação e análises, estes pares estão dispostos na Tabela 5. Os valores para o parâmetro L foram selecionados com a mesma justificativa usada na primeira aplicação, compreendendo a periodicidade de 24, então são múltiplos de 24. Os valores para r foram escolhidos de acordo com gráfico para cada L .

Tabela 5 - Valores de L e r testados para a segunda aplicação

L	r
72	35
96	47
120	63
144	53
240	75
360	83

O par de parâmetros que melhor se ajustou a série em questão foi 120 e 63. As análises feitas para chegar a esse resultado foram as mesmas da aplicação anterior. Também foi obtido de forma visual utilizando o gráfico de autovalores, matriz de correlações, e gráfico de autovetores pareados, analisados de forma conjunta. Confirmados com os gráficos das funções de autocorrelação e da autocorrelação parcial (FAC e FACP).

Para o gráfico de autovalores e matriz correlação w (Figuras 21 e 22, respectivamente), nota-se que não há uma visibilidade que comprove o melhor valor a ser utilizado por r para que aconteça a melhor separação entre as componentes. No gráfico de autovalores (Figura 21) é possível observar que há um decaimento a partir da componente 40, o que pode significar uma divisão entre as componentes sinal e ruído naquele intervalo.

Autovalores L=120

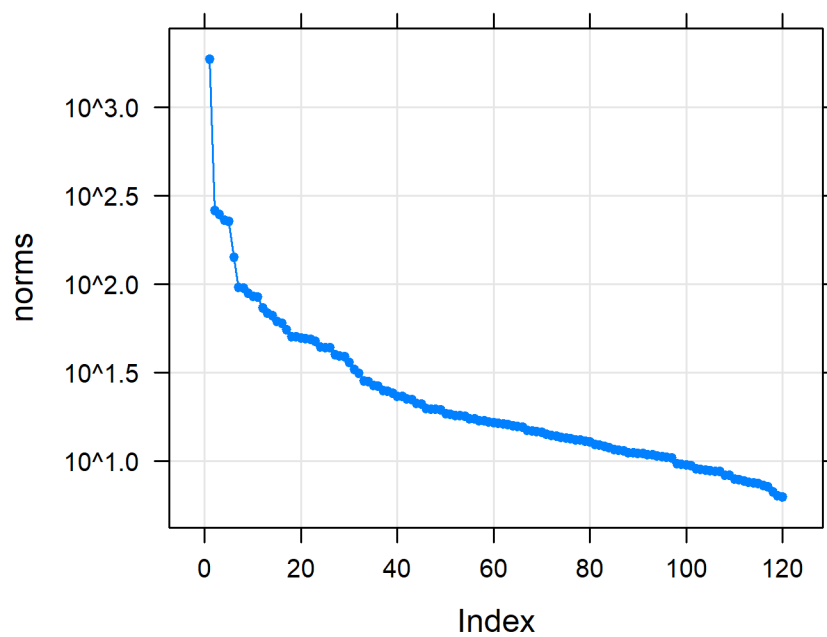


Figura 21 - Gráfico de autovalores para L = 120

Fonte: Autora (2020)

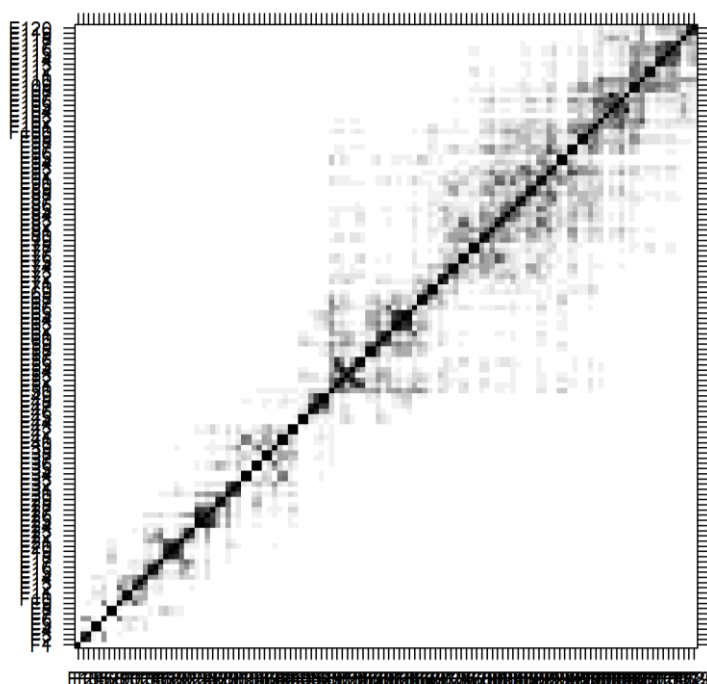


Figura 22 - Matriz de correlações para L = 120

Fonte: Autora (2020)

Se tratando da matriz de correlações a visualização se torna mais difícil, não sendo possível identificar de forma clara essa separabilidade, devido a grande quantidade de componentes essa identificação se torna inviável através dessa ferramenta. Deste modo, uma última análise a ser feita para se obter um número para r é através da análise do gráfico dos autovetores pareados que apresenta uma visão mais ampla da relação entre as componentes.

De acordo com o gráfico dos autovetores pareados, o número de autotriplas ideal para esse conjunto de dados a ser utilizado nessa reconstrução é $r = 63$, como pode ser visto na Figura 23, onde a partir deste ponto, nenhuma imagem geométrica pode ser formada, ou visivelmente mencionada, o que fica caracterizada como componentes de ruído.

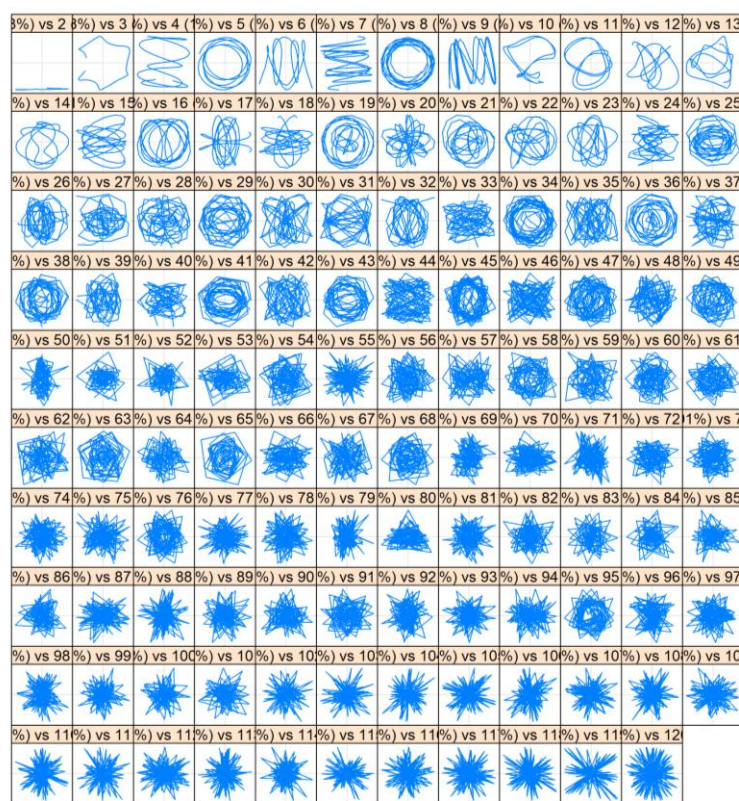


Figura 23 - Gráfico de autovetores pareados para $L = 120$

Fonte: Autora (2020)

Então, foram utilizados esses dois parâmetros para a decomposição e reconstrução da série e obtenção da série mais próxima da original, e posterior previsão seguindo o método em questão.

Do mesmo modo como foi feito para série com os dados anuais, para a série em questão também foi feita a análise sobre os ruídos aplicando o SSA no ruído para confirmação de que o mesmo só havia componente ruído, e análise da FAC

e FACP, que confirmaram que na série havia pouca dependência de ruído. Com isso verifica-se que foi feita uma boa separação das componentes.

O próximo passo foi a previsão e comparação com os diferentes tamanhos de janela e números de autotriplas, como pode ser visto na Tabela 6 através do RMSE e do SMAPE, e no gráfico representado na Figura 24 para as previsões das 24 horas do dia 31/12. A variação dos parâmetros ajuda a ratificar o que já foi encontrado nas análises anteriores, ou ter outras conclusões.

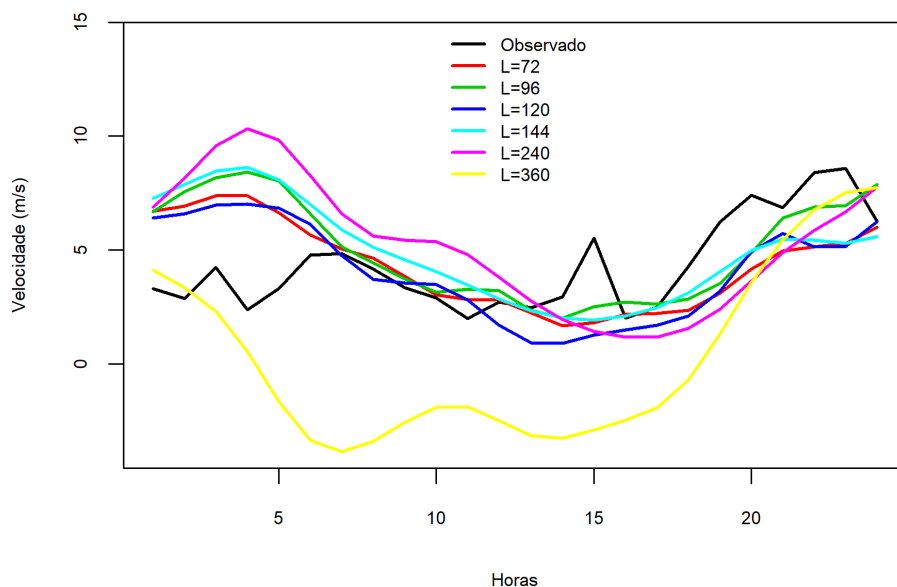


Figura 24 - Previsões 24 horas para os tamanhos de janela para o dia 31/12 na segunda aplicação

Fonte: Autora (2020)

Além de uma análise visual também é importante observar os valores encontrados para as métricas para cada par de parâmetro, esses resultados estão apresentados na Tabela 6, sendo a comparação entre as previsões feitas com as combinações dos valores L e r através das métricas RMSE e SMAPE. O menor valor obtido de acordo com o RMSE foi para $L = 120$ e para a métrica SMAPE foi $L = 96$. Havendo essa divergência entre os melhores valores encontrados a partir das métricas, escolheu-se utilizar o valor encontrado de acordo com o RMSE.

Tabela 6 - Métricas para os parâmetros na segunda aplicação

$L - r$	RMSE	SMAPE
72 - 35	2,420	40,05%
96 - 47	2,479	37,71%
120 - 63	2,400	49,26%
144 - 53	2,739	42,24%
240 - 75	3,443	61,08%
360 - 83	4,937	137,52%

Observa-se que todas as janelas acompanharam de forma favorável os dados observados, com exceção de $L = 360$ que teve um comportamento mais “distorcido” em relação as outras, apresentando um RMSE mais elevado e um SMAPE bem maior do que os outros pares analisados. No entanto, as demais apresentaram bons desempenhos quando comparados com os dados reais.

4.2.2.1

Análise de sensibilidade do tamanho de janela para a segunda aplicação

Os mesmos pares testados de tamanhos de janelas (L) e número de autotriplas (r) utilizados para esta segunda aplicação, foram estão utilizados para prever diversos dias do mês de dezembro, mudando somente o conjunto de treinamento. Deste modo, é possível visualizar como as janelas selecionadas adequam-se no decorrer dos últimos dias de dezembro e a sua performance diante das previsões encontradas. A Tabela 7, dispõe todas as janelas com seus respectivos RMSE's e SMAPE's encontrados, e facilita a visualização das janelas que melhores se adequaram aos dias analisados. Vale ressaltar que, todas as janelas em questão, obtiveram bons resultados, considerando os valores das métricas bem próximos para os dias em questão, com pouca oscilação.

Tabela 7 - Melhor par de parâmetros de acordo com o dia previsto

Treinamento	Previsão	$L - r$	RMSE	SMAPE
01/12 – 24/12	25/12	360 - 83	0,808	12,19%
01/12 – 25/12	26/12	96 - 47	1,069	14,80%
01/12 – 26/12	27/12	120 - 63	1,286	16,75%
01/12 – 27/12	28/12	96 - 47	1,838	27,51%
01/12 – 28/12	29/12	240 - 75	0,883	15,31%
01/12 – 29/12	30/12	144 - 53	2,502	44,78%
01/12 – 30/12	31/12	120 - 63	2,400	49,26%

Janelas como 120 e 96 se repetiram, mas no geral, todas se adequaram a algum dia específico, apresentando bons desempenhos no intervalo selecionado.

4.3 Comparação entre as aplicações

A comparação realizada entre as duas aplicações foi realizada através da análise dos melhores desempenhos entre as previsões de acordo com as métricas RMSE e SMAPE para o dia previsto. Primeiramente, foi analisado a métrica RMSE para comparar as duas aplicações, como pode ser visto na Tabela 8, em seguida, na Figura 25 pode ser analisada visualmente essa comparação.

Tabela 8 - Valores do RMSE para as duas aplicações

Dia previsto	RMSE (Ano todo)	RMSE (Dezembro)
25/12	1,301	0,808
26/12	1,026	1,069
27/12	1,158	1,286
28/12	1,366	1,838
29/12	0,76	0,883
30/12	2,719	2,502
31/12	1,705	2,400
Média	1,434	1,541

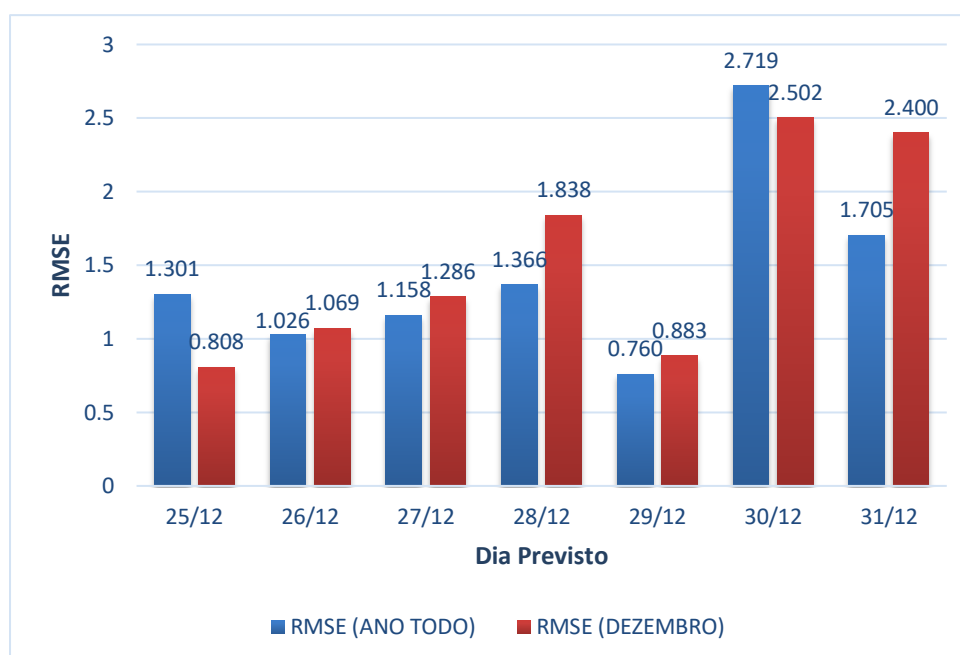


Figura 25 - Comparação entre o desempenho das previsões com RMSE

Fonte: Autora (2020)

De acordo com os resultados obtidos através da comparação entre os RMSE's (Figura 25), observa-se que dentre os sete dias previstos, cinco deles obtêm uma melhor previsão se utilizado o conjunto de treinamento o ano todo, isto é, se o conjunto conter um número maior de dados. No entanto, não é descartada a viabilidade de ser utilizado somente o último mês do ano para esta situação, visto que para diversos dias as métricas obtiveram valores muito próximos um do outro, principalmente quando observa-se a média entre os métodos.

Para a métrica SMAPE, pode ser observado os valores encontrados para as duas aplicações na Tabela 9, seguindo da comparação feita de forma visual através do gráfico exposto na Figura 26.

Tabela 9 - Valores do SMAPE para as duas aplicações

Dia previsto	SMAPE (Ano todo)	SMAPE (Dezembro)
25/12	18,36%	12,19%
26/12	14,73%	14,80%
27/12	15,27%	16,75%
28/12	22,03%	27,51%
29/12	14,55%	15,31%
30/12	49,30%	44,78%
31/12	31,96%	49,26%
Média	23,74%	25,80%

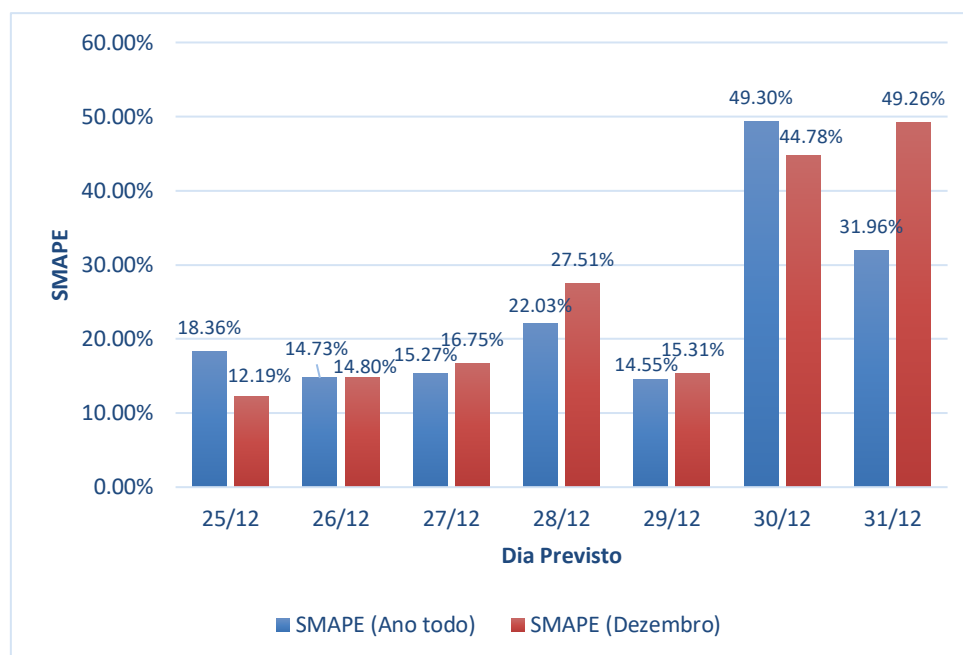


Figura 26 - Comparação entre o desempenho das previsões com SMAPE

Fonte: Autora (2020)

Para a métrica SMAPE os resultados obtidos foram parecidos com os resultados apresentados através da comparação entre os RMSE's (Figura 25). Observa-se que dentre os sete dias previstos, cinco deles obtêm uma melhor previsão se utilizado o conjunto de treinamento o ano todo. No entanto, não é descartada a viabilidade de ser utilizado somente o último mês do ano para esta situação, pois em vários dias as métricas obtiveram valores muito próximos sendo possível considerar essa opção.

Em suma, diante do que foi apresentado, o modelo utilizando o conjunto de dados do ano obteve um bom desempenho diante do conjunto de dados quando analisado os dias previstos.

No entanto, é possível observar que apesar da maioria dos dias terem uma previsão maior utilizando o conjunto de dados do ano todo, caso seja feita uma análise hora a hora pode-se obter melhores resultados quando utilizado o conjunto do mês de dezembro. Esse comportamento é visível quando analisadas as duas métricas, mas tomando como exemplo utilizando o RMSE, pode ser visto na Figura 27 o dia 31 de Dezembro onde obteve para o dia inteiro o RMSE melhor se utilizado o ano todo como conjunto de dados. No entanto, algumas horas deste mesmo dia o RMSE foi melhor se utilizado o conjunto do mês de dezembro.

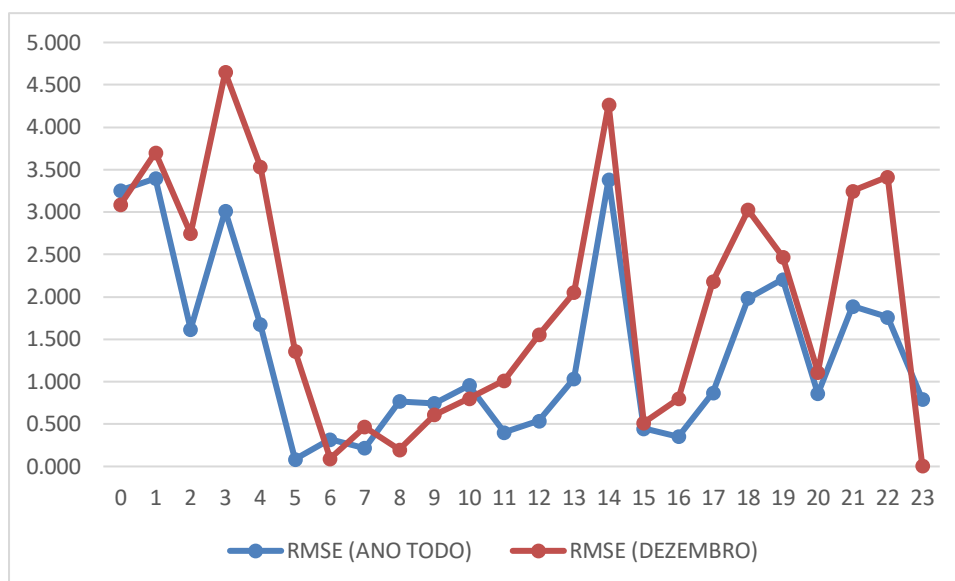


Figura 27 - Variação do RMSE no dia 31/12

Fonte: Autora (2020)

Na Figura 28, acontece uma situação parecida com a supracitada, porém nesse caso o RMSE foi menor quando utilizado os dados do mês de dezembro. Entretanto, quando analisado hora a hora do dia previsto, observa-se que em algumas horas o RMSE utilizando o ano todo teve um bom desempenho, obtendo valores menores.

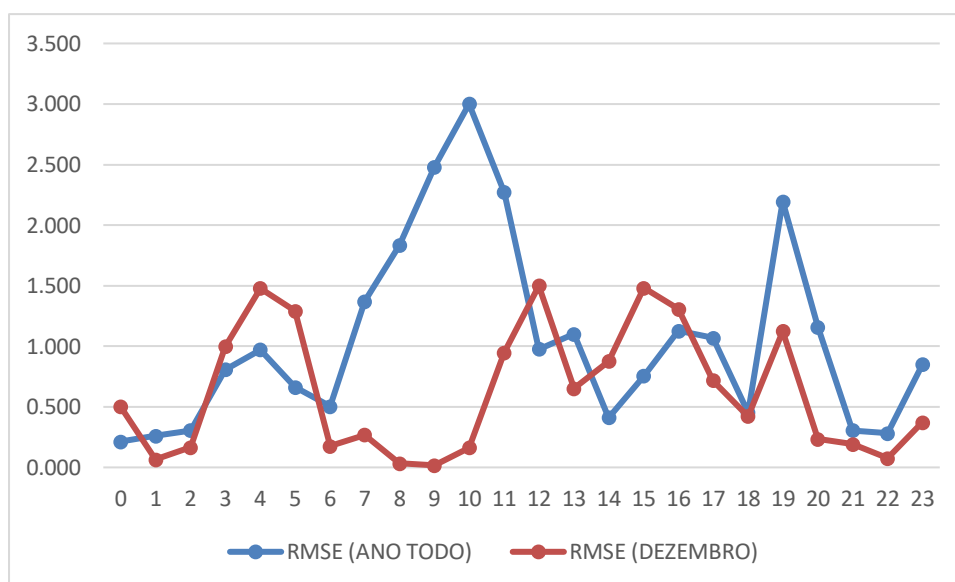


Figura 28 - Variação do RMSE no dia 25/12

Fonte: Autora (2020)

5

Conclusões e pesquisas futuras

A energia eólica é considerada uma das fontes mais atraentes de energia renovável devido à sua alta eficiência e baixa poluição. Deste modo, esse tipo de energia atualmente tem um papel fundamental estrategicamente auxiliando as demais fontes de energia em todo o mundo e vem ganhando espaço nas matrizes energéticas de diversos países. No Brasil, seu uso começou a ser estudado com mais ênfase e implementado após a crise de energia que aconteceu em 2001, conhecida como “apagão”. Depois desse acontecimento, vários investimentos foram feitos nessa área, além de diversos estudos buscando uma inserção maior dessa fonte na matriz energética nacional.

A intermitência do principal recurso utilizado pela fonte eólica, o vento, pode atrapalhar o nível de penetração da mesma e afetar a operação e a confiabilidade. A solução para este problema é uma previsão confiável e precisa, além de ser mais econômico e eficaz. Previsões principalmente de curto prazo são mais úteis para fins operacionais, como aumento ou redução da carga, ou planejamento econômico de despacho de carga.

O estudo teve como objetivo contribuir para a melhoria da previsão da velocidade do vento para geração de energia eólica no período de curto prazo, através da aplicação da técnica de previsão *Singular Spectrum Analysis (SSA)* a uma usina eólica localizada no Brasil, fazendo uma análise comparativa de modelos SSA considerando diferentes horizontes de previsão e conjunto de treinamento, para diferentes dias de previsão, com diferentes tamanhos de série temporal. Sendo assim, uma comparação da série temporal de um ano todo foi feita com somente o último mês desta série, prevendo os últimos sete dias do mês de dezembro.

Diante das aplicações realizadas, o estudo atingiu o objetivo proposto aplicando a técnica primeiramente em um conjunto maior de dados (ano todo) e em seguida utilizando somente o último mês (Dezembro). Depois as devidas comparações foram realizadas utilizando as métricas RMSE e SMAPE. De modo a analisar qual modelo melhor se adequa a situação proposta.

Quando são utilizados dados de alta frequência o desafio é ainda maior, principalmente porque a técnica exige a determinação dos parâmetros tamanho de janela e número de autotriplas, o que para dados com uma frequência menor já é algo bem trabalhoso, no caso analisado foi um obstáculo, por se tratar de dados horários sendo observados ao longo de um ano completo. No entanto, o estudo realizado pode ser visto como uma contribuição acadêmica visto que pouco se tem visto relacionado a aplicações com dados de alta frequência.

Os resultados foram positivos, no entanto muitos estudos ainda devem ser realizados de forma a explorar mais profundamente esse campo de pesquisa. Por ser uma técnica ainda pouco aplicada, principalmente voltada para dados de alta frequência, o SSA ainda necessita de maiores análises para aplicações desse tipo, no entanto, a técnica apresentou resultados promissores.

Diante disso, conclui-se com este estudo que a relação *L e r* varia muito dependendo do intervalo estudado, tamanho da série e histórico, deste modo, não sendo possível afirmar a existência da relação entre esses dois parâmetros quando se abrangem dias diferentes.

Como pesquisas futuras propõe-se mais estudos e outras análises mais profundas devem ser feitas utilizando dados de alta frequência na aplicação da técnica SSA, de forma a obter resultados mais conclusivos a respeito, tendo em vista que este campo pode apresentar resultados promissores com estudos mais aprofundados.

Referências bibliográficas

AASIM; SINGH, S. N.; MOHAPATRA, A. Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting. **Renewable Energy**, v. 136, p. 758–768, 2019.

ABEEÓLICA. Números ABEEólica. p. 9, 2019.

AGUILAR VARGAS, S. Previsão da distribuição da densidade de probabilidade da Geração de Energia Eólica usando técnicas não paramétricas. Tese de Doutorado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2015.

BENEKI, C.; EECKELS, B.; LEON, C. Signal extraction and forecasting of the UK tourism income time series: A singular spectrum analysis approach. **Journal of Forecasting**, v. 31, n. 5, p. 391–400, 2012.

CADENAS, E. et al. Wind speed prediction using a univariate ARIMA model and a multivariate NARX model. **Energies**, v. 9, n. 2, p. 1–15, 2016.

CADENAS, E.; JARAMILLO, O. A.; RIVERA, W. Analysis and forecasting of wind velocity in chetumal, quintana roo, using the single exponential smoothing method. **Renewable Energy**, v. 35, n. 5, p. 925–930, 2010.

CADENAS, E.; RIVERA, W. Short term wind speed forecasting in La Venta, Oaxaca, México, using artificial neural networks. **Renewable Energy**, v. 34, n. 1, p. 274–278, 2009.

CASSIANO, K. M. Análise de Séries Temporais Usando Análise Espectral Singular (SSA) e Clusterização de Suas Componentes Baseada em Densidade. Tese de Doutorado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2014.

CATALÃO, J. P. S.; POUSINHO, H. M. I.; MENDES, V. M. F. Short-term wind power forecasting in Portugal by neural networks and wavelet transform. **Renewable Energy**, v. 36, n. 4, p. 1245–1251, 2011.

CHEN, Q.; FOLLY, K. A. Wind Power Forecasting. **IFAC-PapersOnLine**, v. 51, n. 28, p. 414–419, 2018.

DE CARVALHO, M.; RODRIGUES, P. C.; RUA, A. Tracking the US business cycle with a singular spectrum analysis. **Economics Letters**, v. 114, n. 1, p. 32–35, 2012.

DE GIORGI, M. G. et al. Comparison between wind power prediction models based on wavelet decomposition with Least-Squares Support Vector Machine (LS-SVM) and Artificial Neural Network (ANN). **Energies**, v. 7, n. 8, p. 5251–5272, 2014.

DOWELL, J.; PINSON, P. Very-Short-Term Probabilistic Wind Power Forecasts by Sparse Vector Autoregression. **IEEE Transactions on Smart Grid**, v. 7, n. 2, p. 763–770, 2016.

EISSA, M. et al. Wind Power Prediction Using a Hybrid Approach with Correction Strategy Based on Risk Evaluation. **INTERNATIONAL JOURNAL of RENEWABLE ENERGY RESEARCH** M.Eissa et al, v. 7, n. 3, 2017.

EMPRESA DE PESQUISA ENERGÉTICA (EPE). **Fontes**. Disponível em: <<http://www.epe.gov.br/pt/areas-de-atuacao/energia-eletrica/expansao-da-geracao/fontes>>. Acesso em: 25 maio. 2019.

ERDEM, E.; SHI, J. ARMA based approaches for forecasting the tuple of wind speed and direction. **Applied Energy**, v. 88, n. 4, p. 1405–1414, 2011.

FILIK, T. Improved spatio-temporal linear models for very short-term wind speed forecasting. **Energies**, v. 9, n. 3, 2016.

GANI, A. et al. Degradation of lignin in ionic liquid with HCl as catalyst. **Environmental Progress & Sustainable Energy**, v. 35, n. 3, p. 809–814, 2015.

GOLYANDINA, N.; NEKRUTKIN, V.; ZHIGLJAVSKY, A. Analysis of Time Series Structure: SSA and Related Techniques. Chapman & Hall, 2001.

GUO, Z. et al. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. **Renewable Energy**, v. 37, n. 1, p. 241–249, 2012.

GWEC. Global Wind Statistics 2017. Global Wind Energy Council. 2018.

HASSANI, H. et al. On the separability between signal and noise in singular spectrum analysis. **Fluctuation and Noise Letters**, v. 11, n. 2, 2012.

HASSANI, H. et al. Forecasting accuracy evaluation of tourist arrivals. **Annals of Tourism Research**, v. 63, p. 112–127, 2017.

HASSANI, H. et al. Monthly forecasting of GDP with mixed-frequency multivariate singular spectrum analysis. **International Journal of Forecasting**, v. 35, n. 4, p. 1263–1272, 2019.

HASSANI, H.; HERAVI, S.; ZHIGLJAVSKY, A. Forecasting European industrial production with singular spectrum analysis. **International Journal of Forecasting**, v. 25, n. 1, p. 103–118, 2009.

HASSANI, H.; SOOFI, A. S.; ZHIGLJAVSKY, A. A. Predicting daily exchange rate with singular spectrum analysis. **Nonlinear Analysis: Real World Applications**, v. 11, n. 3, p. 2023–2034, 2010.

LANDBERG, L. Short-term prediction of local wind conditions. **Journal of Wind Engineering and Industrial Aerodynamics**, v. 89, n. 3–4, p. 235–245, 2001.

LIU, H.; MI, X.; LI, Y. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM. **Energy Conversion and Management**, v. 159, n. December 2017, p. 54–64, 2018.

LIU, R.; PENG, M.; XIAO, X. Ultra-Short-Term Wind Power Prediction Based on Multivariate Phase Space Reconstruction and Multivariate Linear Regression. **Energies**, v. 11, n. 10, p. 2763, 2018.

LÓPEZ-MANRIQUE, L. M. et al. Assessment of resource and forecast modeling of wind speed through an evolutionary programming approach for the north of tehuantepec isthmus (Cuauhtemotzin, Mexico). **Energies**, v. 11, n. 11, p. 1–22, 2018.

MA, X.; JIN, Y.; DONG, Q. A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized by brain storm optimization for short-term wind speed forecasting. **Applied Soft Computing Journal**, v. 54, p. 296–312, 2017.

MAJUMDER, S.; KANJILAL, P. P. Application of Singular Spectrum Analysis for Investigating Chaos in Sea Surface Temperature. **Pure and Applied Geophysics**, v. 176, n. 8, p. 3769–3786, 2019.

MORENO, S. R.; DOS SANTOS COELHO, L. Wind speed forecasting approach based on Singular Spectrum Analysis and Adaptive Neuro Fuzzy Inference System. **Renewable Energy**, v. 126, p. 736–754, 2018.

NANDANA JYOTHI, M.; RAMANA RAO, P. V. Very-short term wind power forecasting through wavelet based ANFIS. **International Journal of Power Electronics and Drive Systems**, v. 9, n. 1, p. 397–405, 2018.

QIAN, Z. et al. A review and discussion of decomposition-based hybrid models for wind energy forecasting applications. **Applied Energy**, v. 235, n. April 2018, p. 939–953, 2019.

RODRIGUES, P. C.; MAHMOUDVAND, R. The benefits of multivariate singular

spectrum analysis over the univariate version. **Journal of the Franklin Institute**, v. 355, n. 1, p. 544–564, 2018.

RODRIGUES, P. C.; TUY, P. G. S. E.; MAHMOUDVAND, R. Randomized singular spectrum analysis for long time series. **Journal of Statistical Computation and Simulation**, v. 88, n. 10, p. 1921–1935, 2018.

SOMAN, S. S. et al. A review of wind power and wind speed forecasting methods with different time horizons. **North American Power Symposium 2010, NAPS 2010**, p. 1–8, 2010.

TIAN, Z.; REN, Y.; WANG, G. Short-term wind speed prediction based on improved PSO algorithm optimized EM-ELM. **Energy Sources, Part A: Recovery, Utilization and Environmental Effects**, v. 41, n. 1, p. 26–46, 2019.

TRONCOSO, A. et al. Local models-based regression trees for very short-term wind speed prediction. **Renewable Energy**, v. 81, p. 589–598, 2015.

WANG, C.; ZHANG, H.; MA, P. Wind power forecasting based on singular spectrum analysis and a new hybrid Laguerre neural network. **Applied Energy**, v. 259, n. August 2019, p. 114139, 2020.

YANG, Z.; WANG, J. Multistep wind speed forecasting using a novel model hybridizing singular spectrum analysis, modified intelligent optimization, and rolling elman neural network. **Mathematical Problems in Engineering**, v. 2016, 2016.

YU, C.; LI, Y.; ZHANG, M. An improved Wavelet Transform using Singular Spectrum Analysis for wind speed forecasting based on Elman Neural Network. **Energy Conversion and Management**, v. 148, p. 895–904, 2017.

ZHANG, W. et al. Hybrid wind speed forecasting model study based on SSA and intelligent optimized algorithm. **Abstract and Applied Analysis**, v. 2014, 2014.

ZHAO, E. et al. Hybrid wind speed prediction based on a self-adaptive ARIMAX model with an exogenous WRF simulation. **Energies**, v. 9, n. 1, p. 1–20, 2016.

ZHAO, Y. et al. Correlation-Constrained and Sparsity-Controlled Vector Autoregressive Model for Spatio-Temporal Wind Power Forecasting. **IEEE Transactions on Power Systems**, v. 33, n. 5, p. 5029–5040, 2018.

ZHU, X.; GENTON, M. G. Short-Term Wind Speed Forecasting for Power System Operations. **International Statistical Review**, v. 80, n. 1, p. 2–23, 2012.