

**Universidade Federal de Pernambuco
Centro de Tecnologia de Geociências
Departamento de Engenharia Elétrica
Programa de Pós-Graduação em Engenharia Elétrica**

HELEN BARBOZA SILVA

**TÉCNICAS PARA REDUÇÃO DE DIMENSIONALIDADE DE SÉRIES TEMPORAIS
E DETECÇÃO DE VELOCIDADES EXTREMAS DO VENTO PARA GERAÇÃO
EÓLICA**

RECIFE

2018

HELEN BARBOZA SILVA

**TÉCNICAS PARA REDUÇÃO DE DIMENSIONALIDADE DE SÉRIES TEMPORAIS
E DETECÇÃO DE VELOCIDADES EXTREMAS DO VENTO PARA GERAÇÃO
EÓLICA**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Pernambuco como parte dos requisitos para obtenção do grau de Mestre em Engenharia Elétrica.

Área de Concentração: Processamento de Energia.

Orientador: Prof. Dr. Ronaldo Ribeiro Barbosa de Aquino

Coorientador: Prof. Ph.D. Manuel Herrera Fernandez

RECIFE

2018

Catálogo na fonte
Bibliotecária Maria Luiza de Moura Ferreira, CRB-4 / 1469

S586t Silva, Helen Barboza.
Técnicas para redução de dimensionalidade de séries temporais e detecção de velocidades extremas do vento para geração eólica / Helen Barboza Silva. - 2018.
80 folhas, il., tab., abr., sigl. e simb.

Orientador: Prof. Dr. Ronaldo Ribeiro Barbosa de Aquino.
Coorientador: Prof. Ph.D. Manuel Herrera Fernandez.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG.
Departamento de Engenharia Elétrica, 2018.
Inclui Referências.

1. Engenharia Elétrica. 2. Regressão quantílica. 3. *Cut-in*. 4. *Cut-out*.
5. Velocidade do vento. 6. Energia eólica. I. Aquino, Ronaldo Ribeiro Barbosa de (Orientador). II. Fernandez, Manuel Herrera (Coorientador). III. Título.

UFPE

621.3 CDD (22. ed.) BCTG/2018-188



Universidade Federal de Pernambuco

Pós-Graduação em Engenharia Elétrica

PARECER DA COMISSÃO EXAMINADORA DE DEFESA DE
DISSERTAÇÃO DO MESTRADO ACADÊMICO DE

HELEN BARBOZA DA SILVA

**“TÉCNICAS PARA REDUÇÃO DE DIMENSIONALIDADE DE SÉRIES
TEMPORAIS E DETECÇÃO DE VELOCIDADES EXTREMAS DO VENTO
PARA GERAÇÃO EÓLICA”**

A comissão examinadora composta pelos professores: RONALDO RIBEIRO BARBOSA DE AQUINO, DEE/UFPE, MANUEL HERRERA FERNANDEZ, ACE-EDEN/UB, GERALDO LEITE TORRES, DEE/UFPE, e ALCIDES CODECEIRA NETO, DEE/UPE, sob a presidência do primeiro, consideram o candidato **HELEN BARBOZA DA SILVA APROVADO.**

Recife, 22 de fevereiro de 2018.

MARCELO CABRAL CAVALCANTI
Coordenador do PPGEE

RONALDO RIBEIRO BARBOSA DE AQUINO
Orientador e Membro Titular Interno

MANUEL HERRERA FERNANDEZ
Coorientador e Membro Titular Externo

GERALDO LEITE TORRES
Membro Titular Interno

ALCIDES CODECEIRA NETO
Membro Titular Externo

Dedico este trabalho as pessoas mais importantes da minha vida, minha mãe, Maria Istra Ferreira Rodrigues da Silva e meu pai Wilson Barboza da Silva (in memória).

AGRADECIMENTOS

Primeiramente a Deus, por todos os propósitos que tem reservado para minha vida e por ter me colocado e recolocado em Seu caminho por tantas vezes.

Agradeço a meus pais, Wilson e Istra, por todos os esforços que fizeram para me criar, dando boa educação, tanto em termos acadêmicos como em termos de caráter. Por terem me dado todo o suporte, compreensão e dedicação que precisei durante minha vida e por todo o amor que me deram.

Agradeço às minhas irmãs, Saskia, Ingrid, Schirley e Suelen, e a toda minha família, por todo apoio e carinho que me deram.

Agradeço a Thalita e Bianca pelo apoio e boas conversas, me mostrando o verdadeiro significado da expressão amiga irmã.

Agradeço também aos meus amigos Angélica, Erik, David, Davidson, Guilherme e Roberto pelo apoio, cafezinhos, caronas e brincadeiras. Vocês fizeram com que estar na universidade todos os dias fosse um prazer.

Agradeço de forma especial a Jonata, pela disponibilidade apresentada durante a elaboração do trabalho, você foi essencial em sua construção.

Agradeço aos meus orientadores Prof. Ronaldo, Prof. Manoel Herrera e Prof. Aida, pela confiança, apoio, direcionamento e supervisão durante a elaboração desse trabalho. Além da amizade desenvolvida.

E enfim, agradeço a todos que de alguma forma passaram em minha vida contribuindo em minha formação.

"O cientista não é o homem que fornece as verdadeiras respostas; é quem faz as verdadeiras perguntas". (Claude Lévi-Strauss)

RESUMO

O alto custo da produção de energia, juntamente com as vantagens da energia eólica como uma fonte de energia renovável e a característica de ser amplamente disponível, levou vários países a estabelecerem incentivos para regular e promover a geração de energia eólica. Todas as turbinas eólicas instaladas no mundo até o final de 2016 podem gerar cerca de 5% da demanda mundial de eletricidade. O Brasil tem um enorme potencial de energia eólica e a diversificação de sua matriz de energia está se tornando cada vez mais necessária, o que é ainda mais relevante quando a fonte de energia é renovável. No mundo, o uso de energia eólica para a produção de energia em larga escala vem recebendo incentivos crescentes, tanto no desenvolvimento de novas tecnologias como em controle e interpretação de informações coletadas para o desenvolvimento de tecnologias nessa área. Uma das mais promissoras soluções para obter informações das séries temporais de velocidade do vento e potência envolvem a realização da redução de dimensionalidade dessas bases de dados; por isso este trabalho propõe a comparação de três métodos: *Piecewise Aggregate Approximation* (PAA), *Piecewise Aggregate Approximation* com Regressão Quantílica (PAA/RQ) e *Adaptive Piecewise Aggregate Approximation* (APAA), para extrair o máximo de informações dessas séries temporais por redução de dimensionalidade, propiciando assim um ganho de tempo computacional no processo de análise de dados. A implementação desses algoritmos permitiu precisar qual o método mais eficiente na determinação da quantidade possível de eventos extremos de *cut-in* e *cut-off* da velocidade do vento e assim determinar quais regiões têm maior potencial para a implementação de um Parque Eólico, assim como estimar períodos em que a turbina eólica não estaria em pleno funcionamento. O desempenho de todos os métodos estudados nesse trabalho é testado para os sítios eólicos localizados nas regiões Sul e Nordeste. Esses locais foram escolhidos por pertencerem a regiões do Brasil com diferentes características geográficas e de vento; essa escolha teve como objetivo a verificação da robustez dos métodos em cenários divergentes. Os resultados indicam que o método APAA teve um desempenho superior que os métodos PAA e PAA/RQ, pois identificou uma maior quantidade de valores extremos.

Palavras-chave: Regressão quantílica. *Cut-in*. *Cut-out*. Velocidade do vento. Energia eólica.

ABSTRACT

The high cost of energy production, coupled with the advantages of wind power as a renewable energy source and the characteristic of being widely available, led several countries to establish incentives to regulate and promote wind power generation. All wind turbines installed in the world by the end of 2016 can generate about 5 % of the world's electricity demand. Brazil has a huge potential for wind energy and the diversification of its energy matrix is becoming increasingly necessary, which is even more relevant when the energy source is renewable. In the world, the use of wind power for large-scale energy production has been receiving increasing incentives, both in the development of new technologies and in the control and interpretation of information collected for the development of technologies in this area. One of the most promising solutions to obtain information on the time series of wind speed and power involves the accomplishment of the dimensionality reduction of these databases, so this work proposes the comparison of three methods: Piecewise Aggregate Approximation (PAA), Aggregate Approximation with Quantile Regression (PAA/QR) and Adaptive Piecewise Aggregate Approximation (APAA), to extract the maximum information from these time series by reducing dimensionality, thus providing a computational time gain in the data analysis process. The implementation of these algorithms made it possible to determine the most efficient method to determine the possible amount of extreme cut-in and cut-off events of the wind speed and thus determine which regions have the greatest potential for the implementation of a wind farm, as well as to estimate periods in which the wind turbine would not be fully operational. The performance of all methods studied in this work is tested for wind sites located in the south and northeast regions. These sites were chosen because they belong to regions of Brazil with different geographic and wind characteristics, this choice had as objective the verification of the robustness of the methods in divergent scenarios. The results indicate that the APAA method performed better than the PAA and PAA/RQ methods because it identified a greater amount of extreme values.

Keywords: Quantile regression. Cut-in. Cut-out. Wind speed. Wind energy.

LISTA DE ILUSTRAÇÕES

Figura 1 – Capacidade Mundial Total Instalada 2001-2016 [MW]	18
Figura 2 – Matriz Elétrica Brasileira 2016.	19
Figura 3 – Capacidade Total Instalada no Brasil.	20
Figura 4 – Formação dos vento a partir do deslocamento das massas de ar.	24
Figura 5 – Representação do escoamento em uma montanha.	27
Figura 6 – Sombreamento da Torre - Cálculo de T_{somb}	28
Figura 7 – Efeito Esteira.	28
Figura 8 – Colocação das turbinas num parque eólico.	29
Figura 9 – Comportamento da função de distribuição de densidade de Weibull para diversos valores de k	31
Figura 10 – Perfil Vertical da Velocidade do vento.	31
Figura 11 – Variação da Velocidade do vento com a altura sobre o terreno, segundo a Lei exponencial de Hellmann.	33
Figura 12 – Representação ilustrativa das componentes de tendência e sazonalidade. . .	38
Figura 13 – Série temporal contínua (A) e discreta (B)	39
Figura 14 – Exemplo de uma matriz e da rota W obtidas a partir da aplicação da DTW. .	44
Figura 15 – Representação ilustrativa da diferença de alinhamento entre as Distâncias Euclidiana e DTW.	44
Figura 16 – Série temporal na representação PAA.	47
Figura 17 – Comparação visual das técnicas de PAA e APAA para reduzir a dimensionalidade de séries temporais.	49
Figura 18 – Ajuste de um Modelo Linear e diversos ajustes da Regressão Quantílica . .	53
Figura 19 – Série temporal para o SERS referente aos anos 2014, 2015 e 2016.	56
Figura 20 – Série temporal para o SERNE referente aos anos 2014, 2015 e 2016.	57
Figura 21 – BoxPlot - Padrões de velocidade do vento para os anos 2014, 2015 e 2016 para o SERS	59
Figura 22 – BoxPlot - Padrões de velocidade do vento para os anos 2014, 2015 e 2016 para o SERNE	59
Figura 23 – Fluxograma com as etapas de cada método aplicado	61
Figura 24 – Representação dos resultados PAA para os valores extremos UP e LOW . .	64
Figura 25 – Representação dos resultados PAA/RQ para os valores extremos de UP e LOW	65
Figura 26 – Representação dos resultados APAA para os valores extremos de UP e LOW	65
Figura 27 – Gráficos de múltiplas barras de ocorrências LOW dos anos de 2014, 2015 e 2016 para SERS com (a) Valores empíricos, (b) Técnica PAA/RQ, (c) Técnica PAA e (d) Técnica APAA	67

Figura 28 – Gráficos de múltiplas barras de ocorrências UP dos anos 2014, 2015 e 2016 para SERS com (a) Valores empíricos, (b) Técnica PAA/RQ, (c) Técnica PAA e (d) Técnica APAA	68
Figura 29 – Gráficos de múltiplas barras de ocorrências LOW dos anos 2014, 2015 e 2016 para SERNE com (a) Valores empíricos, (b) Técnica PAA/RQ, (c) Técnica PAA e (d) Técnica APAA	70

LISTA DE TABELAS

Tabela 1	– Tipos de superfície e comprimentos de rugosidade correspondentes	26
Tabela 2	– Valores do expoente de Hellmann em função da rugosidade do terreno . . .	32
Tabela 3	– Quantidade de janelas para cada conjunto de dados	60
Tabela 4	– Demonstrativo dos valores obtidos de ocorrências LOW e UP para as técnicas PAA/RQ e APAA para o SERS	62
Tabela 5	– Demonstrativo dos valores obtidos de ocorrências LOW e UP para as técnicas PAA/RQ e APAA para o SERS	62
Tabela 6	– Demonstrativo dos valores obtidos de ocorrências LOW e UP para as técnicas PAA/RQ e APAA para o SERNE	63
Tabela 7	– Demonstrativo dos valores obtidos de ocorrências LOW e UP para as técnicas PAA/RQ e APAA para o SERNE	64

LISTA DE ABREVIATURAS E SIGLAS

APAA	<i>Adaptive Piecewise Aggregate Approximation</i>
CPD	<i>Change Point Detection</i> - Detecção de Mudança de Ponto
DFT	<i>Discrete Fourier Transform</i> - Transformada Discreta de Fourier
DTW	<i>Dynamic Time Warping</i>
DWT	<i>Discrete Wavelet Transformation</i> - Transformada Discreta de Wavelet
ED	<i>Euclidean Distance</i> - Distância Euclidiana
GW	Giga Watt
GRP	<i>Glass-Reinforced Polyester</i> - Poliéster Reforçado com Vidro
GWEC	<i>Global Wind Energy Council</i> - Conselho Global de Energia Eólica
HCS	<i>Hill Climb Search</i>
HMM	<i>Global Wind Report</i> - Conselho Global de Energia Eólica
MQO	Mínimos Quadrados Ordinários
MW	Mega Watt
MPPT	<i>Maximum Power Point Tracking</i> - Rastreamento do Ponto de Máxima Potência
PAA	<i>Piecewise Aggregate Approximation</i>
ST	Série Temporal
SEN	Sistema Elétrico Nacional
ONS	Operador Nacional do Sistema Elétrico
PLA	<i>Piecewise Linear Approximation</i> - Aproximação Linear por Partes
PRE	Produção de Energia Renovável
PSF	<i>Power Signal Feedbac</i>
PROINFA	Programa de Incentivo às Fontes Alternativas de Energia Elétrica
RL	Regressão Linear

RLS	Regressão Linera Simples
RQ	Regressão Quantílica
RPM	Rotações por Minuto
RQ	Regressão Quantílica
SAX	<i>Symbolic Aggregate Approximation</i> - Representação Simbólica Aproximada
SVD	<i>Singular Value Decomposition</i> - Decomposição de Valores Singulares
TSR	<i>Tip Speed Ratio</i>

LISTA DE SÍMBOLOS

z_0	Comprimento de Rugosidade
T_{somb}	Torque Aerodinâmico
K_{somb}	Fator de Redução Percentual do Torque Mecânico
$f_{pos}(\theta)$	Torque Aerodinâmico
k	Fator de Forma de Weibull
c	Fator de Escala de Weibull
v	Velocidade do Vento
Γ	Função Gama
v_h	Velocidade do Vento à altura h
v_r	Previsão de Velocidade do Vento a uma altura de referência
h_r	Altura de referência
α	Expoente de Helmann
m/s	Metros por segundo
v_p	Velocidade de Partida do Vento
v_{nom}	Velocidade Nominal
v_c	Velocidade de Corte do Vento
m	Massa do ar
ρ	Massa Volumétrica do ar
A	Área Frontal da Massa de ar
z	Espessura da Massa de ar
E_c	Energia Cinética
P_v	Potência Disponível no Vento
P_{mec}	Potência Mecânica
C_p	Coefficiente de Desempenho da Turbina

β_p	Ângulo de Passo das Pás
S_t	Sazonalidade
γ	Tendência
α_t	Resíduo
μ	Média
σ	Desvio Padrão
L_p	Medidas de Norma
x_{min}	Valores de Mínimo
x_{max}	Valores de Máximo
n'	Dimensão da Série Temporal Reduzida
ε	Erro
v_{min}	Velocidade Mínima
v_{max}	Velocidade Máxima
β_0	Intercepto
β_1	Coefficiente Angular
e_i	Erros Aleatórios
X	Variável Resposta
Y	Variável Preditora
Q_τ	Quantil de ordem τ
$\rho(\tau)$	Termo Ponderador
$\beta_{0i}(\tau)$	Constante da Regressão
$\beta_{1i}(\tau)$	Coefficiente de Regressão
$e_i(\tau)$	Erros Aleatórios Independentes e Identicamente Distribuídos
Y_t	Variável Independente
N	Quantidade de Janelas
m	Quantidade de medições por Mês

SUMÁRIO

1	INTRODUÇÃO	18
1.1	Objetivos	21
1.2	Análise dos Dados	21
1.3	Organização do Trabalho	21
2	PERFIL DO VENTO	23
2.1	Mecanismos de Formação do Vento	23
2.2	Fatores que influenciam na Velocidade do vento	25
2.2.1	Rugosidade do Terreno e Obstáculos	25
2.2.2	Orografia	26
2.2.3	Efeito Sombreamento	27
2.2.4	Efeito Esteira	28
2.3	Modelagem do Vento	29
2.3.1	Distribuição estatística de Weibull	30
2.3.2	Gradiente Vertical de Velocidade do Vento	31
2.3.3	Valores Extremos de Velocidade	33
2.3.4	Energia Útil do Vento	34
3	METODOLOGIA UTILIZADA	36
3.1	Fundamentos das Séries Temporais	36
3.1.1	Pré-Processamento em Séries Temporais	39
3.1.1.1	<i>Valores Faltantes</i>	40
3.1.1.2	<i>Amostragem Irregular</i>	40
3.1.1.3	<i>Normalização</i>	40
3.1.1.4	<i>Normalização min-max</i>	41
3.1.1.5	<i>Z - Normalização</i>	41
3.1.2	Medidas de Similaridade em Séries Temporais	42
3.1.2.1	<i>Distância Euclidiana</i>	42
3.1.2.2	<i>Dynamic Time Warping</i>	43
3.1.3	Representação e Indexação de Séries Temporais	45
3.1.3.1	<i>Piecewise Aggregate Approximation</i>	46
3.1.3.2	<i>Adaptive Piecewise Aggregate Approximation</i>	47
3.1.4	Regressão Quantílica	49
3.1.4.1	<i>Modelo de Regressão Quantílica</i>	52
4	RESULTADOS E DISCUSSÕES	54

4.1	Descrição dos Dados	54
4.2	Imputação de Valores Ausentes	57
4.3	Padrões de Velocidades Analisados	58
4.4	Aplicação dos Métodos Propostos	59
4.5	Resultados	65
5	CONCLUSÃO	71
	REFERÊNCIAS	73

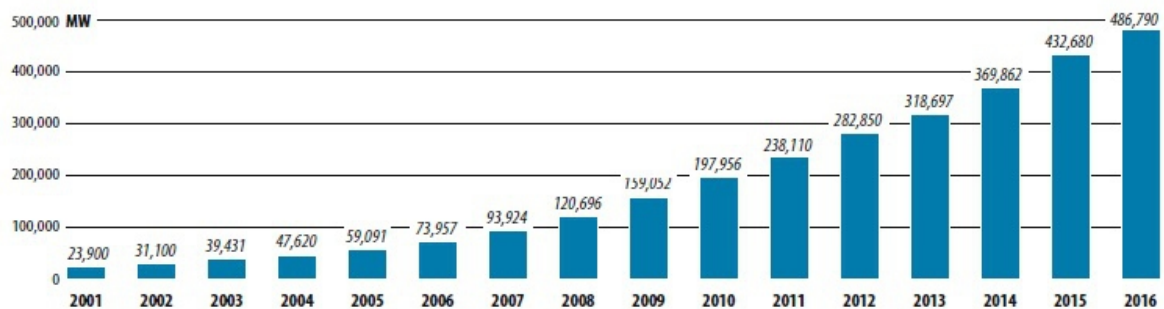
1 INTRODUÇÃO

A produção de energia é uma das principais preocupações do mundo hoje em dia, devido à demanda crescente de energia no mundo. A energia eólica desponta como uma das principais fontes alternativas de energia no mundo, destacando-se pelo reduzido impacto ambiental, por seu avanço tecnológico e por políticas governamentais de incentivo à geração de energia limpa. Além disso, fontes renováveis como eólica e biomassa têm sido cada vez mais consideradas como soluções alternativas para reduzir a dependência das fontes convencionais e complementar o suprimento mundial.

Esse crescimento é liderado pela China, quando ela resolveu expandir seus recursos eólicos com a finalidade de expandir a economia rural, aumentando a capacidade produtiva e estimulando um crescimento sustentável. Sua produção corresponde a quase metade de toda energia eólica gerada no planeta. Segundo GWEC (2016), a Ásia possui a maior parte da geração eólica com 203,685 MW instalados até o final de 2016, seguida do continente Europeu com 161,330 MW. O terceiro continente é a América do Norte com cerca de 88,283 MW, cuja potência instalada vem principalmente dos Estados Unidos. Em seguida vem a América Latina e o Caribe com 12,218 MW, seguidos bem de perto pela Região do Pacífico com 4,823 MW. E por último a África e o Oriente Médio com 3,488 MW até o final de 2016 (GWEC, 2016).

Um relatório do Conselho Global de Energia Eólica revela que a potência total global instalada no final de 2016 no mundo foi de 486,8 MW, representando um crescimento cumulativo do mercado de mais de 12% (GWEC, 2016). Como pode-se perceber no gráfico da Figura 1.

Figura 1 – Capacidade Mundial Total Instalada 2001-2016 [MW]



Fonte: (GWEC, 2016)

No Brasil a maior parte da energia gerada é por meio dos seus recursos hídricos e usinas térmicas, como pode ser observado na Figura 2. Devido a essa alta dependência, o Brasil passou por uma grande escassez de energia em 2001, que afetou exponencialmente a economia do país.

Como forma de controle de situações como essa, em 2002 o governo brasileiro aprovou medidas de incentivo à participação de fontes renováveis de energia na matriz energética

brasileira. Dentre elas aprovou o PROINFA – Programa de Incentivo a Fontes Alternativas de Energia Elétrica, para encorajar a geração de energia por pequenas hidrelétricas, energia eólica e biomassa.

A energia eólica, devido a características como natureza inesgotável, limpa e custo benefício alto, tornou-se uma das fontes de energia renovável, dentro das Produção de Energia Renovável (PRE's), que possui a maior potência instalada no território nacional, onde espera-se que com a sua implementação o Brasil seja incluído entre os 10 países com maior potência instalada de centrais eólicas (LEMOS, 2005).

Figura 2 – Matriz Elétrica Brasileira 2016.



Fonte:(ABEEOLICA, 2017)

Para o Brasil, o ano de 2016 foi importante para a indústria eólica brasileira, com instalações passando o marco de 10 GW. No final de 2016, a capacidade cumulativa de energia eólica do Brasil totalizou 10,74 GW, representando 7% da capacidade de geração nacional. Em 2016, o Brasil adicionou 2.014 MW de nova capacidade de energia eólica, instalando 947 turbinas eólicas em 81 parques e criando 30 mil novos empregos, com múltiplas vantagens para a economia brasileira.

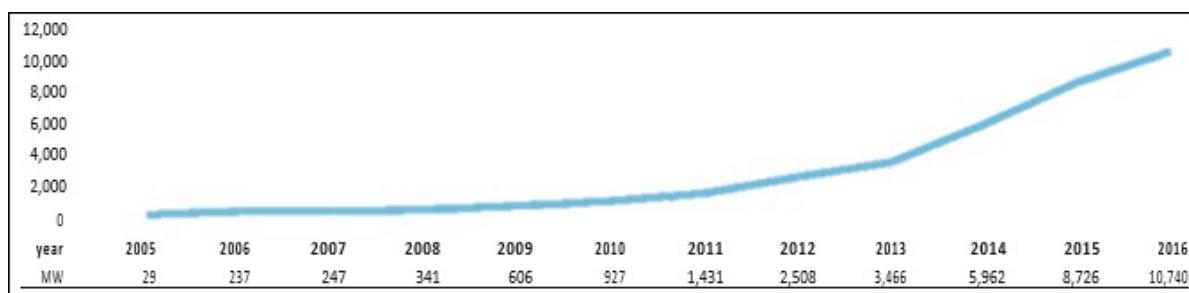
Representando um investimento de US\$ 5,4 bilhões, representando 78% do investimento total em energia limpa no Brasil (GWEC, 2016). A Figura 3 mostra a crescente capacidade total instalada para energia eólica no Brasil do ano de 2005 a 2016.

Devido às incertezas que estão associadas ao vento, mostra-se necessário a utilização de técnicas ou ferramentas que possibilitem da melhor maneira possível, a quantificação da energia que será produzida por estas centrais a curto, médio e longo prazo, propiciando assim uma melhor gestão do Sistema Elétrico Nacional (SEN).

Portanto, a identificação dos valores extremos de velocidade, que correspondem a faixas de velocidades do vento que estão acima de 20 m/s e abaixo de 3 m/s, é de fundamental importância para que assim, se possa determinar os períodos em que a turbina eólica estará em

pleno funcionamento, como também auxilia na definição da área mais propícia à instalação dos parques eólicos.

Figura 3 – Capacidade Total Instalada no Brasil.



Fonte: (ABEEOLICA, 2017)

Considerando o alto valor do investimento necessário para a implantação de parques eólicos, o estudos de técnicas de apoio a tomada dessas decisões se torna essencial, pois propicia uma maior segurança na viabilidade econômica do projeto.

Essas técnicas também possibilitam um equilíbrio na oferta de energia quando associado à geração hidráulica, permitindo uma maior disponibilidade da água acumulada e otimização do uso dos reservatórios, com o aproveitamento desse recurso em períodos secos e em horários de ponta do sistema de segurança dos sistemas elétricos o aspecto de complementação da energia gerada necessária para a carga instalada.

Avanços na área de eletrônica e informática propiciaram melhorias na aquisição de dados e seu armazenamento. Esse avanço fez com que a coleta de dados se tornasse mais eficiente, acarretando estudos mais abrangentes e com altas resoluções temporais para a análise de eventos pontuais. Contudo, a grande quantidade de dados pode inferir em dificuldades na sua manipulação, controle e interpretação; isso pode ocorrer devido a limitações humanas de manipulação de grande quantidade de dados, ou limitações técnicas.

Uma melhoria no método de amostragem foi proposta em (KEOGH et al., 2001), chamado de *Piecewise Aggregate Approximation* (PAA) e sua evolução foi proposta em (CHAKRABARTI et al., 2002), a qual foi chamada de *Adaptive Piecewise Aggregate Approximation* (APAA). Na primeira abordagem é utilizada a média de cada segmento de janelas de tamanho igual para representar o conjunto correspondente de pontos de dados, otimizando a análise da série temporal; já na segunda, a série temporal será representada por segmentos de tamanhos distintos, o que sugere uma maior sensibilidade com relação à localização desses valores extremos nas bases de dados utilizadas, buscando uma análise mais refinada na base de dados.

Para o PAA realizou-se a análise de modelos de regressão quantílica (HERRERA et al., 2016), que é baseado em *quantilis* inferiores e superiores ao invés da regressão ordinária que é condicionada pela média (KEOGH et al., 2001). E para o APAA foi feito o refinamento na análise da base de dados, com a modificação do tamanho de cada segmento da série temporal que poderia conter valores de *cut-in* ou *cut-out*. E ao final desse trabalho conclui-se qual dos

métodos apresentou melhores resultados para a identificação desses valores nas séries temporais de velocidade do vento.

As séries temporais utilizadas neste trabalho consistem em valores de velocidade do vento medidos a cada 10 minutos, começando às 0h de 1 de janeiro de 2014 e terminando às 23h de 31 de dezembro de 2016. Os bancos de dados correspondem a medições obtidas pelo Operador Nacional do Sistema Elétrico (ONS) nas regiões Sul e Nordeste.

1.1 Objetivos

Essa dissertação tem como objetivo a análise e comparação da aplicação de três métodos de redução de dimensionalidade de séries temporais, os algoritmos *Piecewise Aggregate Approximation* (PAA), *Piecewise Aggregate Approximation* com a Regressão Quantílica (PAA/RQ) e *Adaptive Piecewise Aggregate Approximation* (APAA) para análise dos valores extremos dessas séries de vento. Fazendo com que a identificação desses valores seja feito de forma mais eficiente e sem comprometer a sazonalidade dessas séries temporais.

Espera-se que as abordagens desse estudo contribuam de forma significativa para a identificação de regiões com potencial para se tornarem parques eólicos a partir da análise das velocidades extremas do vento, como também a identificação dos períodos em que a turbina eólica estará em pleno funcionamento.

Deseja-se também com esse trabalho que essas técnicas sejam utilizadas como uma opção a IAR, por apresentar características e conceitos mais simples para o entendimento de pessoas leigas na área, propiciando uma maior aplicabilidade em empresas.

1.2 Análise dos Dados

A análise dos resultados será por meio da comparação dos métodos propostos, onde serão avaliados: a quantidade total de valores extremos identificados, estações do ano onde ocorrem a maioria dos eventos. A implementação dessas técnica será realizada por meio da ferramenta MATLAB ®.

1.3 Organização do Trabalho

Esta dissertação está dividida , além desta introdução, em 6 capítulos, ordenados na mesma sequência em que os estudos foram realizados e desenvolvidos. Estes capítulos estão distribuídos conforme é descrito a seguir:

- **Capítulo 1 - Introdução**

É mostrada a importância da energia eólica no Brasil e no mundo, assim como seu potencial eólico atual dentro da matriz elétrica brasileira. Também é dada uma breve explicação dos métodos utilizados na dissertação para a obtenção dos valores extremos de velocidade do vento e por fim os objetivos deste trabalho.

- **Capítulo 2 - Perfil do Vento**

Este capítulo consiste em uma descrição dos fatores que influenciam no mecanismo de formação dos ventos, bem como a descrição de fatores que irão interferir na velocidade do vento. Além disso, conceitos de modelagem do vento, valores extremos de velocidade do vento e energia útil a partir do vento também são abordados.

- **Capítulo 3 - Aspectos Técnicos da Energia Eólica**

Neste capítulo será mostrado o fundamento físico para transformação da energia do vento em energia elétrica, bem como as configurações do rotor e principais componentes de um sistema eólico e suas respectivas funções no processo de geração eólica.

- **Capítulo 4 - Métodos Estudados para Identificação de Valores Extremos de Velocidade do Vento**

Neste capítulo são apresentados os procedimentos, como também a descrição dos métodos utilizados na elaboração desse trabalho. São apresentadas as soluções computacionais que implementa essa metodologia.

- **Capítulo 5 - Análise dos Resultados e Comparação Entre os Métodos**

Neste capítulo são apresentados os resultados obtidos com a implementação dos métodos sugeridos pelo presente trabalho; para tal, optou-se por representar o desempenho de cada método a partir das 4 estações do ano; com base nesses resultados, foram feitas análises comparativas entre os métodos.

- **Capítulo 6 - Conclusões e Trabalhos Futuros**

São apresentadas as conclusões do trabalho, assim como propostas de trabalhos futuros para que se possa dar continuidade às pesquisas nesse contexto.

2 PERFIL DO VENTO

O uso do vento como fonte de energia elétrica teve grande impulso nos anos 90. Isso aconteceu devido a incentivos para pesquisa e desenvolvimento de novas tecnologias de métodos e equipamentos. O conhecimento do comportamento do vento é de suma importância para realizar-se uma análise da viabilidade de um parque eólico; isso é feito a partir da previsão apropriada do potencial eólico no transcorrer dos anos posteriores ao período de medição em campo.

Como se sabe, as variações apresentadas pelo vento são causadas principalmente pelas mudanças climáticas ao longo do dia e estas são definidas como variações diárias, sazonais, anuais e de curta duração.

As variações diárias e sazonais são importantes para estimar o suprimento da demanda do sistema elétrico; já as variações anuais são importantes para estudos da viabilidade da implantação de parques eólicos. Para se determinar o comportamento do vento em um sítio eólico é necessário o registro de velocidade do vento durante longos períodos de tempo. Como esses registros são feitos em curtos períodos de tempo, o volume de dados é muito grande, tornando a extração dessas informações mais demorada e por vezes complexa.

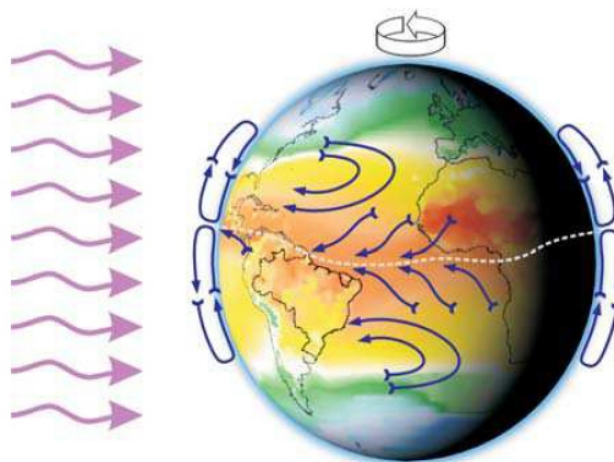
2.1 Mecanismos de Formação do Vento

Os ventos se originam a partir do aquecimento não uniforme da atmosfera, e por isso pode-se considerar que a energia eólica é proveniente da radiação solar. Devido ao movimento da terra e a orientação dos raios solares o sol irá aquecer a terra de modo desigual, fazendo com que as regiões tropicais, mais próximas do sol, sejam mais aquecidas que as regiões polares.

Como resultado, o ar quente que se encontra nas baixas altitudes das regiões tropicais tende a subir, sendo substituído por uma massa de ar mais frio que se desloca das regiões polares; esse deslocamento da massa de ar provoca os fluxos de vento, como pode ser visto na Figura 4 (CARVALHO, 2006),(ABEEOLICA, 2005), (CARVALHO, 2013).

Para a correta utilização de máquinas acionadas pelo vento, é necessário o conhecimento dos tipos de vento, pois existem fatores que modificam o regime geral das máquinas. Apesar da elevada variabilidade, na sua velocidade e direção, existem padrões de circulação global e mecanismos de geração de ventos identificados.

Figura 4 – Formação dos vento a partir do deslocamento das massas de ar.



Fonte: (DUTRA; MONTEZANO; FERREIRA, 2009)

1. **Ventos Planetários ou Constantes:** São aqueles que jamais cessam de “soprar” pois os mecanismos de geração, aquecimento no equador e resfriamento nos pólos, vão estar sempre presentes na natureza. Os ventos planetários ou constantes podem ser classificados em:
 - **Alísios:** ventos que sopram dos trópicos para o equador, em baixas altitudes.
 - **Contra-Alísios:** ventos que sopram do equador para os pólos, em altas altitudes.
 - **Ventos do Oeste:** ventos que sopram dos trópicos para os pólos.
 - **Polares:** ventos frios que sopram dos pólos para as zonas temperadas.

2. **Ventos Continentais ou Periódicos:** São ventos gerados a partir da inclinação do eixo da Terra com relação ao plano da sua órbita em torno do Sol, resultando variações sazonais na intensidade e duração do vento. Os ventos continentais ou periódicos podem ser classificados em:
 - **Brisas:** ventos periódicos que sopram do mar para o continente durante o dia e do continente para o mar durante a noite. Eles são formados devido às diferentes capacidades dessas superfícies de refletir, absorver e eliminar o calor recebido pelo sol. As brisas caracterizam-se por serem ventos periódicos de baixa intensidade, que geralmente não ultrapassam os 50 km/h.
 - **Monções:** ventos periódicos que mudam de direção a cada seis meses aproximadamente. Em escalas de tempo menores que um ano, as monções são previstas com alguma segurança, devido às variações muito lentas, dependendo da estação do ano.

3. **Ventos Locais:** São ventos provenientes de outros mecanismos de formação dos ventos mais específicos, são localizados em determinadas regiões durante um determinado período do ano e possuem características individuais. Existem inúmeros tipos que não se padronizam, uma vez que obedecem às condições naturais de cada localidade, como a pressão atmosférica e as formas de relevo, que direcionam o sentido e destino dos ventos.

2.2 Fatores que influenciam na Velocidade do vento

Segundo CARVALHO (2006), o conhecimento detalhado do regime de vento é de crucial importância tendo em vista que erros na previsão dos ventos conduzem a um mau dimensionamento do sistema e erros na estimativa de produção de energia com consequentes riscos financeiros. Pequenas variações na velocidade do vento causam grandes variações na sua potência em função da relação cúbica entre ambas.

O comportamento estatístico do vento ao longo do dia é um fator que é influenciado pela variação de velocidade do vento ao longo do tempo. Além das características intrínsecas do vento citadas anteriormente a velocidade do vento e sua direção também podem ser influenciadas por fatores externos que atuam sobre elas. No caso da geração eólica existem 4 principais fenômenos que podem causar esses distúrbios, alterando o modo de operação dos aerogeradores. Eles são a rugosidade do terreno, orografia, sombreamento da torre e efeito esteira (CARVALHO, 2006), (FADIGAS; ARLINDO, 2012).

2.2.1 Rugosidade do Terreno e Obstáculos

A rugosidade de um terreno é responsável pela desaceleração do vento próximo ao solo, devido à influência dos obstáculos presentes (casas, depósitos de água, aglomerados de árvores ou mesmo grande estrutura de pedras) e de sua superfície. FELGUEIRAS (2003), fez uma classificação das superfícies quanto à rugosidade, propondo um valor de rugosidade para cada tipo de superfície como é mostrado na Tabela 1.

Tabela 1 – Tipos de superfície e comprimentos de rugosidade correspondentes

Tipo de Superfície	Comprimento de Rugosidade z_0(cm)
Água	~0,001
Areia	0,01 - 0,1
Neve	0,1 - 0,6
Relva (~0,01m)	0,1 - 1
Terreno Lavrado	2 - 3
Erva Alta	4 - 10
Pinhal	90 - 100
Árvores e algumas casas	20 - 40
Cidade ou grande aglomerado de casas	80 - 120
Centros de grandes cidades	200 - 300
Florestas	50
Cultivado	5
Subúrbio	150

Fonte: (FELGUEIRAS, 2003)

2.2.2 Orografia

A Orografia é um elemento muito importante no que diz respeito à caracterização do escoamento atmosférico de uma região, a partir do estudo das nuances do relevo de uma determinada região. Ela é extremamente dependente da complexidade do terreno analisado. As montanhas, depressões e vales podem afetar de forma relevante o escoamento atmosférico de diversas formas.

Quando uma massa de ar encontra uma encosta durante o dia se eleva, entrando em contato com o ar frio durante a noite, causando seu arrefecimento, fazendo com que o ar adjacente à montanha eleve a temperatura ou se arrefeça por condução e mistura (CUSTÓDIO, 2012). Essa diferença de temperatura entre o ar da montanha e o circundante favorece a ocorrência dos fenômenos térmicos.

Quando a estabilidade estática é neutra, o escoamento sobre as montanhas cria gradientes de pressão na direção do escoamento que podem originar a separação da camada limite. Este fenômeno vai provocar turbilhões em esteira a montante ou a jusante da montanha, podendo atingir distâncias de várias vezes a altura da montanha.

Quando a separação ocorre a jusante da montanha, a esteira produzida pode ir até 10 ou 20 vezes a altura do ponto mais alto da montanha, enquanto que, quando ocorre a montante, os turbilhões produzidos “aninham-se” na base da mesma e são "constrangidos" pela sua presença.

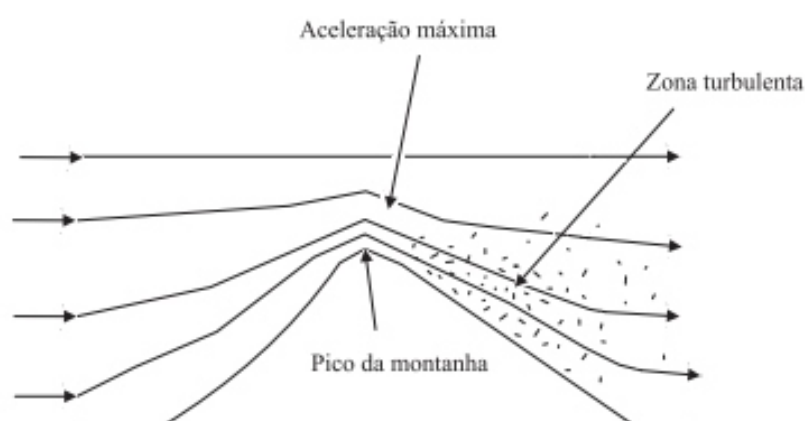
Neste caso, a esteira não vai além de duas alturas da montanha. Este fenômeno é quase inexistente em elevações de inclinação suave, e neste caso, existindo concentração das linhas de corrente, verifica-se um aumento da velocidade do vento e conseqüente aumento de energia que pode ir até 2 vezes o valor obtido sem a perturbação do terreno (FELGUEIRAS, 2003).

Na Figura 5 pode-se notar que o aerogerador terá seu desempenho afetado quando o

escoamento ultrapassa a montanha e é forçado a atingir uma determinada inclinação, contando que esses aerogeradores tenham sido colocados na mesma orientação do escoamento.

Como os aerogeradores são projetados para aproveitamento de componentes horizontais do escoamento, a ocorrência de um declive do fluxo de ar superior ao limite de segurança estabelecido pela norma IEC-61400-1, pode afetar na durabilidade e funcionamento desses equipamentos.

Figura 5 – Representação do escoamento em uma montanha.



Fonte:(FELGUEIRAS, 2003)

2.2.3 Efeito Sombreamento

Em turbinas eólicas de alta potência ocorre um fenômeno que tem impacto direto na qualidade de energia, conhecido como sombreamento da torre, *Tower Shadow*. O sombreamento da torre é o efeito que a própria estrutura causa na velocidade do vento, por se tornar um obstáculo. Porém, este efeito é apenas traduzido em um menor aproveitamento energético quando alguma das pás da turbina se posiciona à frente da torre. Durante um ciclo de uma turbina, esse efeito ocorre com a mesma incidência que o número de pás daquela unidade eólica. Este fenômeno possui uma modelagem senoidal (CUSTÓDIO, 2012), e é definido pela Equação 1.

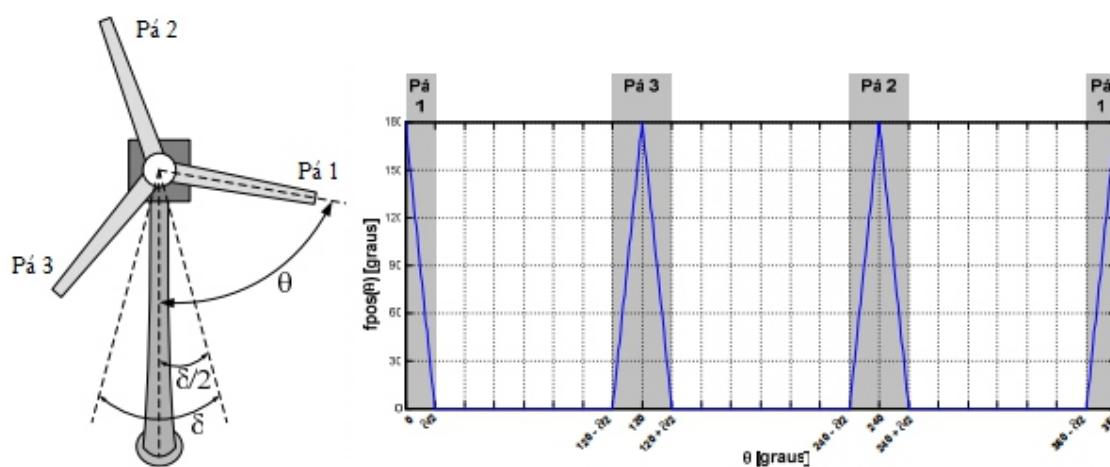
$$T_{somb} = K_{somb} * \left(\frac{1}{2} \cos(f_{pos}(\theta)) - \frac{1}{2} \right) * T_t, \quad (1)$$

em que,

T_{somb} : torque aerodinâmico devido ao sombreamento da torre;

K_{somb} : fator de redução percentual do torque mecânico por influência do sombreamento aerodinâmico;

$f_{pos}(\theta)$: função que assume os valores mostrados na Figura 6, conforme a posição θ das pás da turbina.

Figura 6 – Sombreamento da Torre - Cálculo de T_{somb} .

Fonte:(MACCARINI, 2009)

2.2.4 Efeito Esteira

A turbina eólica produz energia mecânica a partir da energia do vento incidente, podendo haver a diminuição da velocidade do vento ao passar pelas turbinas, já que parte de sua energia é absorvida pelo movimento giratório das pás. E assim, irá se formar uma esteira de vento turbulento com velocidade relativamente reduzida ao vento incidente após as torres. Na Figura 7 pode-se verificar o efeito de esteira obtido a partir da inserção fumo branco no ar que passa através da turbina.

Figura 7 – Efeito Esteira.

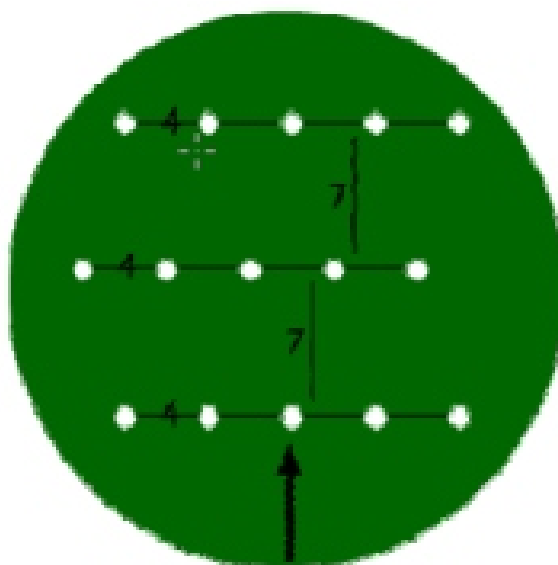


Fonte:(DANISH..., 2010)

Devido ao efeito esteira, a colocação das turbinas dentro de um parque eólico deve ser

realizado de forma cautelosa, como é mostrado na Figura 8. É habitual espaçar as turbinas de uma distância entre cinco e nove diâmetros na direção preferencial do vento e entre três e cinco diâmetros na direção perpendicular. Mesmo tomando estas medidas, a experiência mostra que a energia perdida devido ao efeito de esteira é cerca de 5%.

Figura 8 – Colocação das turbinas num parque eólico.



Fonte:(DANISH... , 2010)

2.3 Modelagem do Vento

O comportamento do vento é modelado a partir de um processo estocástico, tendo a variação aleatória como a velocidade do vento e o índice do processo como o tempo. Devido às variações sazonais do vento, a velocidade média e o desvio padrão não são constantes ao longo de um período, que pode ser um dia, um mês ou até mesmo uma estação do ano. Portanto, a rigor, o vento não é um processo estacionário (COROTIS; SIGI; KLEIN, 1978). Contudo, se as medições forem realizadas de modo a não seguir uma tendência específica de um período qualquer ou tiver uma quantidade relevante, englobando um período longo como anos, este efeito pode ser desprezado (THOMANN; BARFIELD, 1998).

A simulação do vento é importante para a análise estrutural de turbinas eólicas devido à não linearidade entre a turbulência atmosférica e os carregamentos aerodinâmicos. Para a realização destes cálculos, é necessário o uso de séries temporais do vento que simulem seu comportamento turbulento ao longo do tempo, em diferentes pontos da turbina. Para isso é importante apresentar algumas características inerentes ao comportamento dessas séries de velocidade do vento.

2.3.1 Distribuição estatística de Weibull

A Função de Distribuição Estatística de Weibull é usada para representar matematicamente a função densidade de probabilidade de ocorrências das velocidades do vento de uma determinada região. As primeiras aplicações da distribuição de Weibull em estudos eólicos se resumiram à representação da distribuição de velocidades em análises dos carregamentos de vento (DAVENPORT, 1963).

Posteriormente, constatou-se que para a maioria das ocorrências de vento a distribuição de Weibull fornece um bom ajuste (JUSTUS; HARGRAVES; YALCIN, 1976), (HENNESSEY, 1976), (ROHATGI; NELSON, 1994).

A distribuição de Weibull caracteriza-se por dois parâmetros: um de escala, c (em m/s), e outro de forma, k (adimensional) (JUSTUS; HARGRAVES; GRADER, 1978), (TROEN; PETERSEN, 1989), (ARAO, 1989). A função densidade da probabilidade de ocorrência de uma velocidade v é representada matematicamente por (SANSIGOLO, 2005), (JUSTUS; MIKHAIL, 1976):

$$f(v) = \frac{k}{v} \left(\frac{v}{c}\right)^{k-1} \exp \left[-\left(\frac{v}{c}\right)^k \right], \quad (2)$$

em que,

k : fator de forma de Weibull;

c : fator de escala de Weibull ;

v : valor de velocidade do vento.

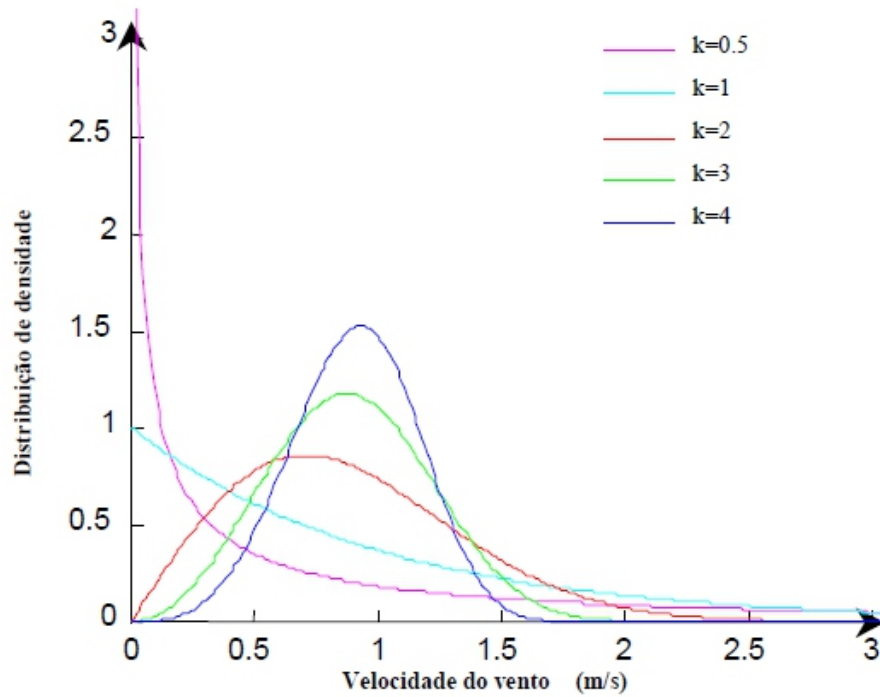
Quanto maior o valor de k , mais estreita e pontuda é a curva de Weibull, indicando uma distribuição mais ou menos uniforme, apresentando uma concentração de ocorrência em torno da média. Já o fator de escala está relacionado ao valor médio da velocidade, não é usado diretamente, sendo determinado através da velocidade média; quanto maior c maior o valor da velocidade média V_{med} , como pode ser observado na Figura 9.

$$V_{med} = c \Gamma \left(1 + \frac{1}{k} \right), \quad (3)$$

em que a função Γ é dada por:

$$\Gamma(y) = \int_0^{\infty} e^{-x} x^{y-1} dx \quad (4)$$

Figura 9 – Comportamento da função de distribuição de densidade de Weibull para diversos valores de k .

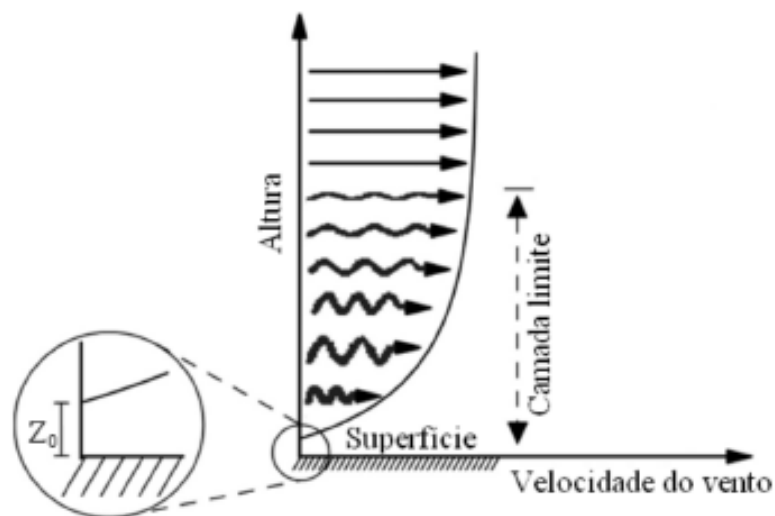


Fonte:(HENNESSEY, 1976)

2.3.2 Gradiente Vertical de Velocidade do Vento

O gradiente de velocidade do vento é determinado pela variação de sua velocidade em relação a altura medida à partir do solo. Como pode ser observado na Figura 10.

Figura 10 – Perfil Vertical da Velocidade do vento.



Fonte:(FADIGAS; ARLINDO, 2012)

O perfil vertical do vento vai depender, segundo (FADIGAS; ARLINDO, 2012) da Lei exponencial de Hellmann, que representa o modelo mais simples e mostra a relação entre a velocidade do vento e a altura de medição. É um parâmetro importante para a escolha da altura do *Hub* dos aerogeradores e indicador dos esforços aerodinâmicos exercidos pelo vento ao longo das pás.

A lei de Hellman é dada pela seguinte equação:

$$v_h = v_r \left(\frac{h}{h_r} \right)^\alpha, \quad (5)$$

em que,

v_h : é a velocidade do vento a altura h ;

v_r : é a previsão de velocidade do vento a uma altura de referência;

h_r : é a altura de referência;

α : é o expoente de Hellmann que varia com a rugosidade do terreno;

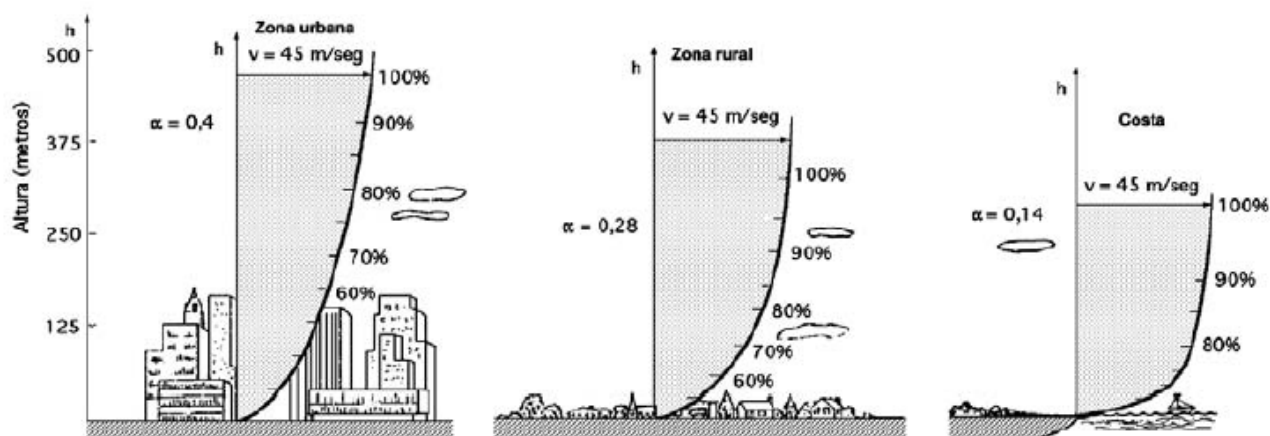
A Tabela 2 mostra os valores do coeficiente exponencial de Hellmann em função da rugosidade do terreno. Na Figura 11, são indicados as variações de velocidade do vento com a altura, segundo a Lei exponencial de Hellmann.

Tabela 2 – Valores do expoente de Hellmann em função da rugosidade do terreno

Tipo de Superfície	Expoente de Hellmann α
Lugares planos com gelo	0,08 - 0,12
Lugares planos (mar, costa)	0,13 - 0,16
Terrenos pouco acidentados	0,2 - 0,26
Zonas rústicas	0,25 - 0,4
Terrenos acidentados ou bosques	0,2
Terrenos muito acidentados ou cidades	0,25 - 0,4

(CUSTÓDIO, 2012)

Figura 11 – Variação da Velocidade do vento com a altura sobre o terreno, segundo a Lei exponencial de Hellmann.



Fonte:(DIEZ, 2003)

2.3.3 Valores Extremos de Velocidade

A percentagem de energia realmente capturada pela turbina eólica depende do fato dela estar à plena operação o tempo todo. Contudo a justificativa para isto não acontecer, é o fato que não existem ventos suficientes em parques eólicos para gerar a capacidade nominal da turbina eólica, ou então esses ventos não estão dentro da faixa considerada segura, a nível estrutural, para funcionamento da turbina.

As velocidades do vento que incidem nas turbinas eólicas devem estar dentro de intervalos de máximo e mínimo, *cut-out* e *cut-in* respectivamente, para estas entrarem e saírem de operação. A geração de energia elétrica a partir do vento inicia-se com velocidades de partida do vento v_p , conhecida como *cut-in wind speed* que é da ordem de 2,5 – 3 m/s (WIND. . . , 2014). As velocidades abaixo desses valores tem sua capacidade energética bem reduzida, pois o torque exercido pelos ventos nas pás é insuficiente para fazer a turbina partir, não fundamentando o aproveitamento. No entanto, na medida que a velocidade vai aumentando a turbina começará a girar e gerar energia elétrica.

A velocidade nominal do vento v_{nom} , *rate wind speed*, é a velocidade na qual a turbina eólica irá alcançar a sua potência nominal, isto é velocidades superiores a 12 m/s e inferiores a 17 m/s ativam o sistema automático de limitação da potência da máquina, que pode ser por controle de ângulo de passo das pás ou por estol aerodinâmico, dependendo do modelo da turbina eólica. Quando a velocidade do vento está entre v_p e v_{nom} a potência gerada cresce com o cubo da velocidade.

A velocidade de corte do vento v_c , conhecida como *cut-out wind speed* que é da ordem de 20 m/s, é a máxima velocidade de trabalho da turbina eólica, que não vai gerar danos a sua estrutura. Nesses casos extremos, um sistema de frenagem é empregado para levar o rotor a uma paralisação ou uma mudança de ângulo em relação à direção dos ventos predominantes.

2.3.4 Energia Útil do Vento

Como já foi visto, o vento é formado a partir do aquecimento irregular do globo. Quando uma parcela de ar se encontra em movimento, uma certa quantidade de energia cinética é associada a ela; assim pode-se dizer que a energia cinética de uma massa de ar m em movimento e a velocidade v , é dada por (FADIGAS; ARLINDO, 2012):

$$E_c = \frac{1}{2}mv^2, \quad (6)$$

$$m = \rho Az, \quad (7)$$

em que,

m : é a massa do ar;

v : é a velocidade de transporte da massa do ar;

ρ : é a massa volumétrica do ar;

A : Área frontal da massa de ar;

z : Espessura da massa de ar.

A potência disponível no vento P_v , é dada pela expressão da energia cinética, derivando a espessura da massa de ar em ordem ao tempo, como mostra a Equação 8 (CARVALHO, 2006),(NETO, 2009), (MACCARINI, 2009).

$$P_v = \frac{dE_c}{dt} = \frac{1}{2}\rho Av^3. \quad (8)$$

A quantidade de energia que o vento possui vai variar de acordo com o cubo da velocidade média do vento, contudo a potência possível de extrair do vento não pode ser totalmente convertida em potência mecânica, pois isto teria como consequência a ausência de qualquer vento após a passagem de ar pela turbina eólica. Desta forma, o máximo da energia cinética do vento, que pode ser convertido para energia mecânica por uma turbina eólica, é determinado pela *Lei de Betz*; ela diz que só se pode converter menos de 16/27 (59%) da energia cinética em energia mecânica usando um aerogerador.

Foi formulada pela primeira vez pelo físico alemão Albert Betz em 1919. Seu livro *Wind-Energy*, publicado em 1926, proporciona boa parte do conhecimento desta época sobre energia eólica e aerogeradores. (BURTON; SHARPE; BOSSANYI, 2001). Assim, a fração de potencia proveniente do vento que é absorvida pela turbina varia a partir do coeficiente de potência C_p dado pela Equação 9:

$$C_p = \frac{P_{mec}}{P_v}. \quad (9)$$

Pode-se expressar a potência mecânica em termos das grandezas do vento pelas Equações 7, 8 e 9. Sendo representada pela Equação 10

$$P_{mec} = \frac{1}{2} \rho A C_p v^3. \quad (10)$$

Sendo, P_{mec} a potência eólica aproveitável e o coeficiente C_p o desempenho da turbina eólica, tendo como máximo valor teórico $16/27$ ou 59% (valor conhecido como "Limite de Betz"). Este coeficiente de potência depende de duas outras variáveis: λ , que relaciona a velocidade do vento com a velocidade da ponta das pás da turbina, e β_p , que corresponde ao ângulo de passo das pás da turbina (BONELLI, 2010).

3 METODOLOGIA UTILIZADA

Para extrair informações das séries temporais é necessário realizar a análise dessas séries através da observância de certas características intrínsecas a ela. Pois a representação de uma série temporal não é determinada pela utilização de apenas um dado ou de uma subsequência isolada. E ainda, a representação de uma série temporal ou subsequência de forma compacta pode acarretar perda ou imprecisão das informações extraídas, caso não sejam utilizadas as técnicas corretas.

Assim, a representação adequada da série de maneira que facilite a extração de conhecimento e que torne fácil a sua manipulação computacional e, ainda, que preserve o máximo das informações originais constituem um dos pilares para a análise de séries temporais (BARIONI, 2006).

Neste capítulo, são apontados conceitos aplicados ao pré-processamento de séries temporais, medidas de similaridade, assim como técnicas de análise das séries temporais para extração do máximo de informações, a partir da redução de dimensionalidade dessas séries, sem comprometer a acurácia e precisão das análises e com menor complexidade computacional.

3.1 Fundamentos das Séries Temporais

Uma série temporal pode ser definida como uma sequência ordenada de valores ou eventos obtidos sobre repetidas medidas ao longo do tempo (HAN; KAMBER, 2006). Uma série temporal de tamanho n pode ser representada por um vetor $\vec{x} = [x_i]$, com $i = 1, 2, 3, \dots, n$. E cada componente x_i é um número real, representando o valor de uma variável no instante i .

A sua unidade de medida irá variar de acordo com a origem dos dados, podendo ser em segundos, minutos, meses, horas ou dias por exemplo. Neste trabalho, considerou-se as unidades de tempo em minutos, ou seja, as séries de vento foram obtidas a cada 10 minutos, então para cada dia do mês foram realizadas 144 medições.

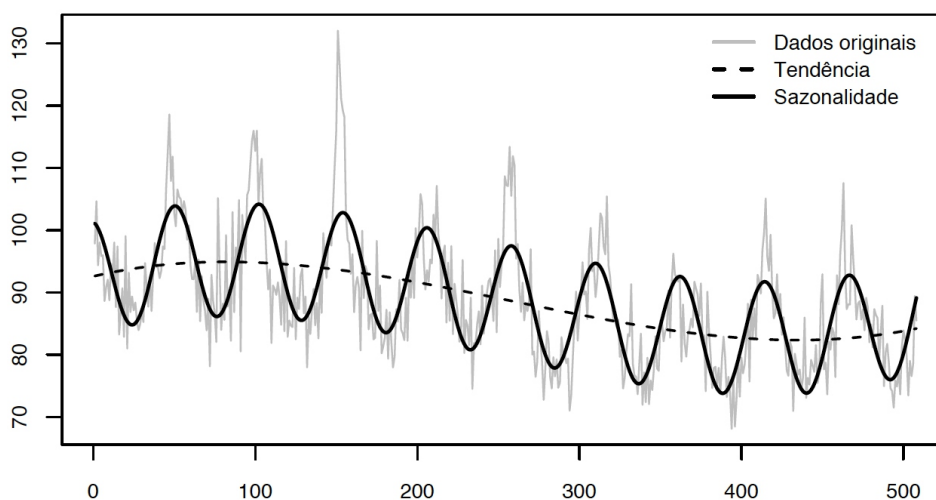
Existem várias dificuldades para a manipulação e análise das séries temporais (KEOGH; KASETTY, 2003), dentre elas:

- **Alta Dimensionalidade:** Uma das particularidades das séries temporais é sua grande dimensionalidade. Quanto mais longa a série, mais informações será possível extrair, contudo sua análise se torna mais difícil e demorada.
- **Subjetividade:** A extração da informação irá depender do usuário, do domínio e da tarefa em questão.
- **Diversidade:** Elas podem ser encontradas em diferentes formatos, registradas em diferentes taxas de amostragens, com ruído, valores ausentes, etc. Essa característica dificulta ainda mais sua análise.

É comum na literatura, decompor a série temporal em um conjunto finito de componentes independentes, para que se tenha uma percepção e análise mais assertiva dos eventos que estas representam. As principais componentes observadas são (MORETTIN; TOLOI, 2004):

- **Sazonalidade:** Ela consiste em comportamentos que tendem a ser repetidos em uma ST em diferentes períodos de tempo. A sazonalidade irá apresentar oscilações ao longo da componente de tendência a partir de uma determinada característica (FERRERO, 2009). Essa componente é de grande importância, pois pode influenciar na identificação de comportamentos relevantes de determinado fenômeno observado ou dificultar a identificação de outros. Assim, a percepção dessa componente é importante no contexto de análise da ST por revelar informações relevantes ou com a sua remoção pode-se ressaltar outras características (MALETZKE, 2014). Essa componente pode ser observada pela linha contínua em preto da Figura 12.
- **Tendência:** Ela corresponde ao movimento dominante em uma ST, influenciando em observações por longos períodos de tempo. Sua característica principal é o movimento quase constante de crescimento ou decrescimento, que atua de modo suave ao longo da série. Ela pode ser observada na Figura 12, a qual é representada pela linha tracejada. As séries temporais podem possuir distintos tipos de comportamentos de tendência, nos quais os métodos de identificação dessa componente estão baseados (EHLERS, 2005). Os métodos mais utilizados nesse tipo de tarefa são: suavizar os valores da série próximos a um determinado ponto, ajustar uma função aos valores observados da série e suavizar os valores da série por meio de sucessivos ajustes das retas de mínimos quadrados ponderados (MORETTIN; TOLOI, 2004).
- **Resíduo:** Ele representa os eventos aleatórios que caracterizam o próprio evento que está sendo analisado, eles são ocasionados pelos equipamentos e procedimentos utilizados durante o período de amostragem. A sua existência não pode ser captada pelas componentes de sazonalidade e tendência, o que dificulta ainda mais a identificação da mesma. Desse modo, a modelagem das demais componentes permite isolar e analisar o resíduo (EHLERS, 2005).

Figura 12 – Representação ilustrativa das componentes de tendência e sazonalidade.



Fonte: (FERRERO, 2009)

Assim uma série temporal T_i composta por $i = 1, \dots, n$ observações é descrita como a soma dessas três componentes como pode ser observado na Equação 11 (MORETTIN; TOLOI, 2004).

$$T_i = \gamma_i S_i a_i, \quad (11)$$

Em que,

S_i : é a sazonalidade;

γ_i : é a tendência;

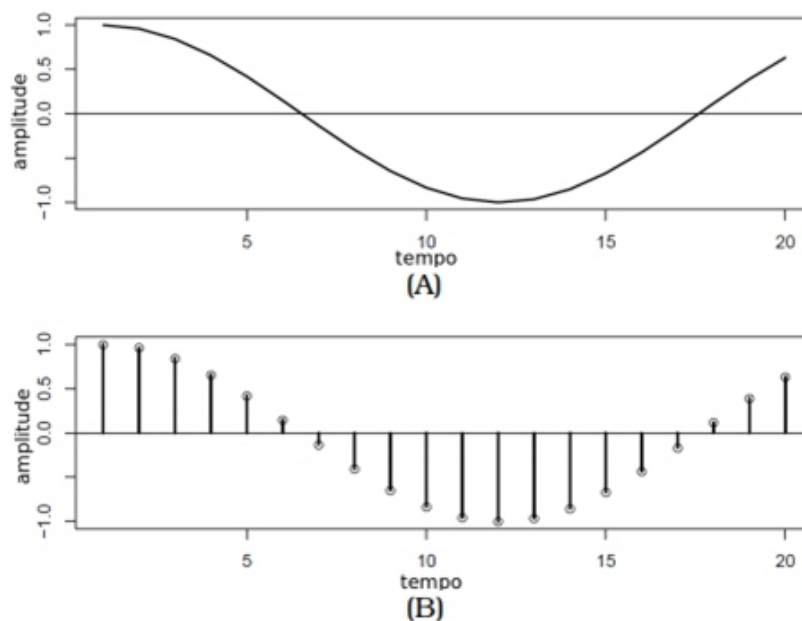
a_i : é o resíduo.

Daí, com a análise das componentes e características das ST é possível fazer uma análise do conteúdo destas tendo como objetivos:

- Compreender o mecanismo da série, possibilitando encontrar razões para o comportamento dela;
- Obter controle dos processos que geram as observações, garantindo que a série tenha um comportamento já esperado;
- Descrever a série mostrando as propriedades constitutivas dela como tendência, sazonalidade entre outras;
- Predizer valores futuros, utilizando dados de comportamentos passados e também métodos de previsão.

Quando as observações são realizadas continuamente no tempo, a série é dita contínua, Figura 13 (A). Quando as observações são realizadas em intervalos de tempo equidistantes, ela é dita discreta, como mostrado na Figura 13 (B), na qual é utilizada uma série temporal hipotética de amplitude $[-1 : 0]$ e comprimento $[0 : 20]$.

Figura 13 – Série temporal contínua (A) e discreta (B)



Fonte: (RIOS, 2013)

Uma série temporal também pode ser classificada como estacionária e não estacionária. As estacionárias, também conhecidas como séries convergentes, permanecem em equilíbrio em torno de um nível médio constante e estão relacionadas à grande parte da teoria de séries temporais. Suas observações evoluem com média $\mu(t)$ e variância $\sigma(t)$ constantes, refletindo alguma forma de equilíbrio estável, Equação 12, (BOX et al., 2008) e (MORETTIN; TOLOI, 2004).

$$\mu(t) = \mu \quad \sigma^2(t) = \sigma^2 \quad (12)$$

3.1.1 Pré-Processamento em Séries Temporais

O pré-processamento dos dados é uma das etapas mais importantes na análise de séries temporais, podendo corresponder a 80% de todo o processo, pois a qualidade dos dados interfere diretamente na qualidade do conhecimento extraído desses dados, (HAN; KAMBER, 2006) e (MICHALSKI, R. S.; KUBAT, M., 1998).

Essa questão se torna ainda mais complexa quando se trabalha com dados temporais, pois tais dados são altamente sensíveis a transformações e a ordem desses dados constitui uma informação valiosa para sua compreensão. Alguns problemas são bastante similares tanto em dados temporais quanto em dados não temporais e algumas técnicas de pré-processamento possuem diversos fatores em comum, embora alguns problemas do mesmo tipo possam ser tratados de modos diferentes.

Ainda, alguns métodos para análise de séries temporais requerem que essas ST sejam

representadas em outro domínio com o intuito de que algumas características presentes na série temporal sejam melhor analisadas. A seguir são descritos alguns dos problemas tipicamente encontrados em dados de séries temporais.

3.1.1.1 Valores Faltantes

Um dos grandes problemas encontrados na base de dados são os valores faltantes. Existem diversos fatores que podem contribuir para isso, tais como, mal funcionamento de equipamentos, falha humana e condições ambientais por exemplo. Esse problema pode ocorrer em qualquer uma das duas dimensões de uma ST; a primeira diz respeito ao atributo tempo e a segunda ao atributo que representa o fenômeno que está sendo observado.

Uma das formas de contornar esse problema é a partir da aplicação de métodos de previsão das séries temporais, como modelos auto-regressivos (MORETTIN; TOLOI, 2004) e métodos de interpolação (MORCHEN, 2013). Uma outra abordagem bastante aplicada consiste em realizar o cálculo da média aritmética das previsões para que seja usada como substituta do valor ausente.

Em FERRERO (2009) é apresentada uma abordagem para a previsão de ST baseada nesse método, a qual pode ser utilizada com o objetivo de contornar o problema de valores faltantes. Desse modo, o problema de dados faltantes pode ser contornado com a escolha de um desses métodos, sendo sua escolha relacionada ao domínio de aplicação ao qual os dados em questão pertencem (MORCHEN, 2013).

3.1.1.2 Amostragem Irregular

A maioria das observações de um fenômeno são realizadas de forma uniforme, ou seja, são realizadas por meio de incrementos regulares. Diversas técnicas assumem tal uniformidade e apresentam-se ineficientes na presença de deslocamentos não constantes.

Assim, é importante observar na ST esse tipo de característica que envolve a maneira como a ST foi coletada e a necessidade de sua transformação para uma ST equiespaçada. Devido à análise que pretende-se realizar, pode-se decidir por calcular novos valores, ponderando o fator tempo. No entanto, essa decisão poderá ocasionar perdas de informações.

3.1.1.3 Normalização

Existem diferentes maneiras de realizar a comparação entre séries temporais, por meio de medidas baseadas em correlação, ajuste de modelos e compressão (MORCHEN, 2013). Contudo, abordagem tradicional e a mais utilizada consiste, em realizar uma comparação morfológica das ST, baseada principalmente, em medidas da norma L_p .

Para que essa abordagem possa ser realizada, é necessário que algumas observações sobre os dados sejam realizadas, como verificar se as ST a serem comparadas estão em uma

mesma unidade e escala. No processo de normalização, os dados são ajustados de modo que seus valores pertençam a um intervalo especificado.

A normalização permite que os dados sejam ajustados a uma mesma escala permitindo uma comparação direta entre seus valores. A normalização dos dados é uma etapa importante no processamento de dados, sendo amplamente estudada na literatura (KEOGH; KASETTY, 2003). A seguir são apresentadas três abordagens tradicionais de normalização (LIN; KEOGH, 2005).

3.1.1.4 Normalização min-max

Consiste em re-escalar a série temporal em um intervalo específico de $[0, 1]$, conservando o formato original da série temporal. O cálculo dessa normalização é realizado a partir do conhecimento dos valores de mínimo x_{min} e máximo x_{max} dos dados originais. A normalização *min-max* é negativamente influenciada quando a série apresenta valores de *outliers*, com valores extremos de mínimo e máximo. O cálculo dessa normalização pode ser realizado por meio da Equação 13.

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}, \quad (13)$$

Em que,

x_{min} : é o valor de mínimo da ST;

x_{max} : é o valor de máximo da ST.

3.1.1.5 Z - Normalização

Ao aplicar esse método de normalização, os valores da ST são ajustados de modo que a média de seus valores seja nula e o desvio padrão igual a um. Assim como na *min-max*, o formato original da série temporal também é conservado, porém de modo mais similar do que a curva obtida após a normalização *min-max*. Contudo, na z-normalização não existe a necessidade de conhecer os valores extremos de mínimo e máximo. Esse método é útil quando há a presença de *outliers* no conjunto de dados que prejudicariam a normalização *min-max* (HAN; KAMBER; PEI, 2011). Essa normalização pode ser obtida por meio da Equação 14. E devido à característica como robustez na presença de *outliers*, ela foi a escolhida para normalização das bases de dados analisadas.

$$x_{norm} = \frac{x - \mu}{\sigma}, \quad (14)$$

Em que,

μ : é o valor médio da ST;

σ : é o desvio padrão da ST.

3.1.2 Medidas de Similaridade em Séries Temporais

A noção de similaridade é um aspecto fundamental para a análise de séries temporais. Formalmente, uma medida de similaridade pode ser definida como uma função distância que dadas duas séries temporais T_x e T_y como entradas, ela irá retornar um valor numérico, Equação 15.

$$Dist : T_x \times T_y \implies R \quad (15)$$

Uma grande quantidade de medidas de similaridade tem sido proposta na literatura, sendo sua escolha baseada na natureza dos dados a serem analisados, dos requisitos específicos de cada domínio de aplicação, e frequentemente, do tipo de representação adotada (MORCHEN, 2013), (FINK, 2004) e (VLACHOS; GUNOPULOS, 2004). Assim, de modo geral as medidas de similaridade devem prover resultados consistentes com a intuição humana, ressaltar comportamentos globais e locais, abstração de distorções e invariância a transformações.

3.1.2.1 Distância Euclidiana

Dentre as métricas utilizadas na área de séries temporais, a distância Euclidiana é a mais empregada para a comparação de séries temporais (KEOGH; KASETTY, 2002). Essa medida determina qual a distância em linha reta entre dois pontos, os quais pertencem a um espaço de m dimensões, onde m corresponde ao tamanho de uma determinada série temporal. A distância euclidiana entre duas séries temporais quaisquer T_x e T_y de tamanho m é a raiz-quadrada da soma do quadrado da diferença, como mostrado pela Equação 16.

$$dist(T_x, T_y) = \sqrt{\sum_{i=1}^m (T_{xi} - T_{yi})^2} \quad (16)$$

Contudo, essa medida de similaridade apresenta algumas limitações quando aplicada ao contexto de ST, dentre as quais, destacam-se a pouca sensibilidade na comparação de ST que possuem um certo grau de defasagem uma em relação a outra, considerando o eixo temporal, e a necessidade de que ambas as ST devem possuir o mesmo tamanho.

Dentre as suas vantagens tem-se o respeito à desigualdade triangular, ser de fácil implementação e ter complexidade temporal linear em relação ao tamanho da ST (KEOGH; KASETTY, 2002), (KEOGH et al., 2007), (CHAKRABARTI et al., 2002) e (KEOGH; LONARDI; RATANAMAHATANA, 2004).

Mesmo com sua sensibilidade a distorções no eixo temporal, foi demonstrado que na prática a distância euclidiana tem boa performance em relação a outras medidas propostas (KEOGH; KASETTY, 2002). No entanto, é recomendado que ou se aplique um pré-processamento na série temporal para remover tais distorções ou que outra medida mais robusta seja utilizada, por exemplo, a *Dynamic Time Warping*. No presente trabalho será utilizada essa

medida de similaridade pelo fato das séries temporais terem o mesmo tamanho e por ser de fácil implementação.

3.1.2.2 Dynamic Time Warping

A *Dynamic Time Warping* (DTW), descrita em (CHU et al, 2002), é amplamente utilizada no contexto de ST defasadas no eixo do tempo, pois apresenta maior robustez na comparação com outras medidas de similaridade. Outra característica importante é a comparação de séries temporais de tamanhos diferentes. Diferentemente da Distância Euclidiana a DTW permite um melhor ajuste entre duas séries temporais, com o objetivo de amenizar possíveis distorções.

Para determinar a DTW entre duas séries temporais, T_x e T_y de tamanhos m_x e m_y , respectivamente, é necessário alinhar ambas as séries. Para isso é construída uma matriz $m_x \times m_y$, na qual cada elemento (i, j) dessa matriz contém a distância $D(x_i, y_j)$ entre dois pontos x_i e y_j , que geralmente é uma distância euclidiana para se determinar a distância entre dois pontos.

Cada elemento (i, j) da matriz corresponde ao alinhamento entre os pontos x_i e y_j . Essa situação é ilustrada na Figura 14, em que a rota $W = w_1, w_2, \dots, w_k, \dots, w_K$ é um conjunto contínuo de elementos da matriz que define o mapeamento entre T_x e T_y . O k -ésimo elemento de W é definido como:

$$w_k = D(x_i, y_j), \quad \max(m_x, m_y) \leq K < m_x + m_y - 1. \quad (17)$$

A rota geralmente é sujeita a várias restrições:

1. Condições de fronteira: $w_1 = (1, 1)$ e $w_K = (m_x, m_y)$, indica que a rota começa e termina na posição diagonalmente oposta da matriz;
2. Continuidade: dado $w_k = (a, b)$ então $w_{k-1} = (a', b')$, onde $a - a' \leq 1$ e $b - b' \leq 1$. Isso restringe os passos na rota para posições adjacentes (incluindo a diagonal adjacente);
3. Monotonicidade: dado $w_k = (a, b)$ então $w_{k-1} = (a', b')$, onde $a - a' \geq 0$ e $b - b' \geq 0$, restringindo os pontos em W a serem uniformemente espaçados no tempo.

Existem diversas rotas que satisfazem às condições mencionadas acima, contudo busca-se apenas a rota que minimize o custo de alinhamento:

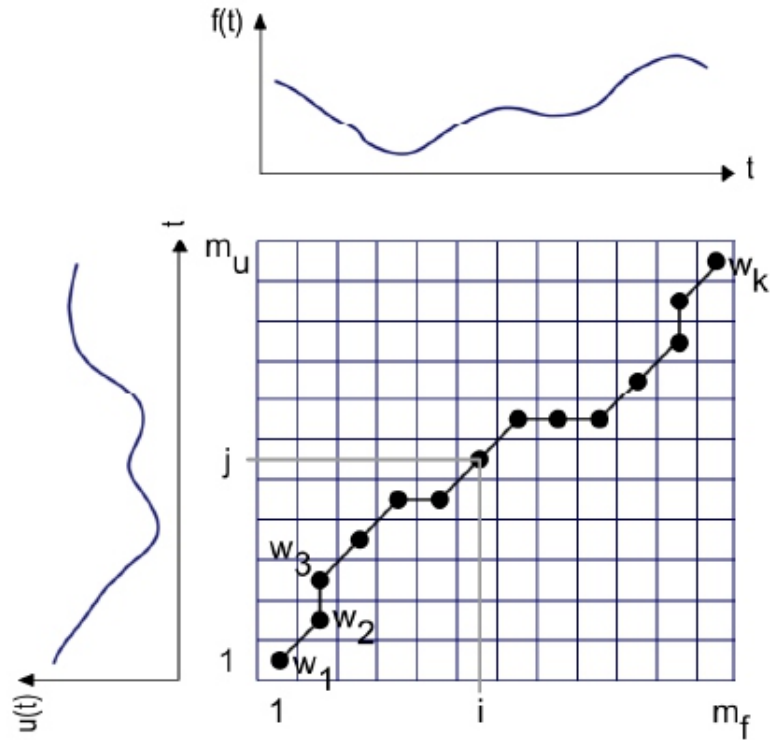
$$DTW(T_x, T_y) = \min \left\{ \frac{1}{K} \sqrt{\sum_{k=1}^K w_k} \right\} \quad (18)$$

A variável K no denominador é usada para compensar o fato de que a rota W pode apresentar diferentes tamanhos. Essa rota pode ser encontrada utilizando programação dinâmica para avaliar as próximas repetições que definem a distância acumulativa $\gamma(i, j)$ como a distância

$D(i, j)$ encontrada na posição atual e a mínima distância acumulativa dos elementos adjacentes, dada pela Equação 19.

$$\gamma(i, j) = D(x_i, y_j) + \min\{\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)\} \quad (19)$$

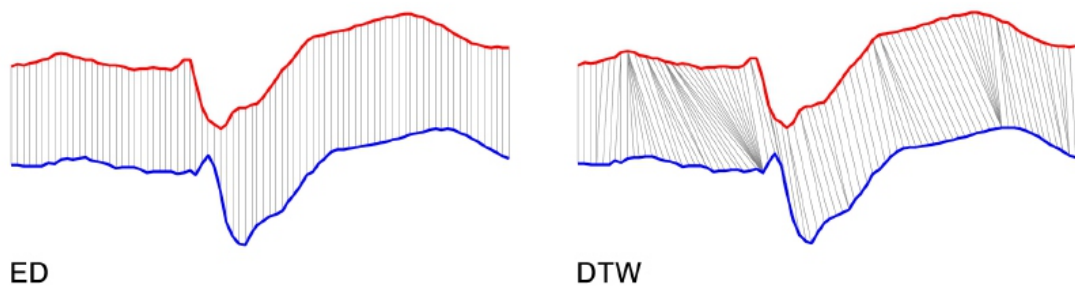
Figura 14 – Exemplo de uma matriz e da rota W obtidas a partir da aplicação da DTW.



Fonte: (CHU et al, 2002)

A Figura 15 representa a diferença de alinhamento entre as medidas de similaridade ED e DTW.

Figura 15 – Representação ilustrativa da diferença de alinhamento entre as Distâncias Euclidiana e DTW.



Fonte: (RATANAMAHATANA; KEOGH, 2004)

3.1.3 Representação e Indexação de Séries Temporais

A alta dimensionalidade observada em ST é proveniente do elevado número de atributos ou conjunto de atributos, o que constitui um sério obstáculo à eficiência da maioria dos algoritmos de extração de conhecimentos de ST. Assim, utilizá-las em seu domínio original torna-se uma tarefa extremamente custosa (HAN; KAMBER; PEI, 2011); para contornar esse problema é desejável representar as séries temporais de modo a simplificá-las, reduzindo sua complexidade, mas ainda assim mantendo suas principais características. A diminuição da dimensionalidade tem variadas vantagens como (CHIZI; MAIMON, 2010):

1. Redução do custo de aprendizagem;
2. Melhoria do desempenho do modelo;
3. Redução das dimensões irrelevantes;
4. Aumento na velocidade computacional
5. Redução das dimensões redundantes.

As séries temporais são na sua maior parte caracterizadas pela sua elevada dimensionalidade, sendo que o tratamento de uma grande quantidade de informação pode ser custoso em termos de processamento e armazenamento; por esse motivo as técnicas de redução de dimensionalidade são desejáveis (WANG et al., 2013). Contudo, estas técnicas devem preservar as características fundamentais do conjunto de dados em análise. A representação de séries temporais pode ser dividida em quatro categorias:

- **Adaptativas:** As técnicas adaptativas utilizam uma representação em comum para todos os elementos da base de dados, de forma a minimizar os erros globais na reconstrução (CHIZI; MAIMON, 2010). Nessa categoria se encontram as técnicas *Singular Value Decomposition* (SVD)(KORN; JAGADISH; FALOUTSOS, 2013), *Symbolic Aggregate Approximation* (SAX)(KORN; JAGADISH; FALOUTSOS, 2013), *Adaptive Piecewise Aggregate Approximation* (APAA) (CHAKRABARTI et al., 2002) e *Piecewise Linear Approximation* (PLA) (CHU et al, 2002).
- **Não-Adaptativas:** As técnicas não adaptativas consideram apenas propriedades locais dos dados e constroem uma representação aproximada da série temporal (WANG et al., 2013). As técnicas mais utilizadas nessa categoria são a Transformada Discreta de Fourier (*Discrete Fourier Transform - DFT*) (AGRAWAL; FALOUTSOS; SWAMI, 1993), a Transformada Discreta de Wavelet (*Discrete Wavelet Transform - DWT*) (CHAN; FU; YU, 2003) e a *Piecewise Aggregate Approximation* (PAA) (CHAKRABARTI et al., 2002).
- **Baseadas em Modelo:** Essas técnicas são fundamentadas em modelos estatísticos ou estocásticos. As ST podem ser representadas a partir de modelos de Markov. Caso eles

possuam parâmetros desconhecidos, eles são denominados de modelos escondidos de Markov (*Hidden Markov Model - HMM*) (RABINER, 1989).

- **Ditadas pelos Dados:** Diferentemente das outras técnicas, as ditadas pelos dados, não necessitam que o usuário determine (implícita ou explicitamente) a taxa de compressão; os dados em si são responsáveis por indicar sua compressão (RATANAMAHATANA et al., 2005). Por exemplo, a técnica conhecida como *clipping* realiza uma discretização binária dos dados, onde uma observação da série temporal é convertida para 0 ou 1, caso seja menor ou maior, respectivamente, que o valor médio da série temporal.

Existem várias técnicas de redução de dimensionalidade. Neste capítulo serão descritas de forma mais aprofundada apenas as técnicas de redução de dimensionalidade de séries temporais utilizadas no presente trabalho, que serão apresentadas nos subitens 4.1.3.1 e 4.1.3.2.

3.1.3.1 Piecewise Aggregate Approximation

A técnica *Piecewise Aggregate Approximation* proposto por LIN et al. (2002), é uma representação que apresenta uma redução significativa da dimensionalidade de uma série temporal sem perdas relevantes de informações e características. A técnica PAA (KEOGH et al., 2001), é uma técnica de redução de dimensão bastante simples de entender e implementar. Ela permite dividir o conjunto de dados das séries temporais em segmentos igualmente espaçados, fornecendo resultados semelhantes ao de transformadas mais elaboradas tais como a Transformada Discreta de Fourier (DFT) e a Transformada Discreta Wavelet (DWT), na tarefa de indexação de séries temporais (LIN et al., 2002).

Esse algoritmo consiste na separação da ST em X segmentos ou janela de tamanho fixo, adjacentes um ao outro, porém disjuntos. Para cada segmento é atribuído um valor que irá representar sua dimensão, sendo que este valor é obtido através do cálculo da média aritmética dos valores que compõem os pontos dentro de cada segmento, obtendo-se uma série formada pelos valores médios de cada um dos segmentos (KEOGH et al., 2001).

Assim, considerando a série temporal X de dimensão n, $X = x_1, \dots, x_n$, define-se como $X' = (x'_1; x'_2; \dots; x'_n)$ a série temporal construída a partir de X, utilizando a técnica PAA, em que x_t representa o valor da t-ésima dimensão, e n' a dimensão da série temporal reduzida. A partir da Equação faz-se o cálculo dos valores da nova representação da série temporal.

$$\bar{x}_t = \frac{n}{n'} \sum_j^i x_j \quad (20)$$

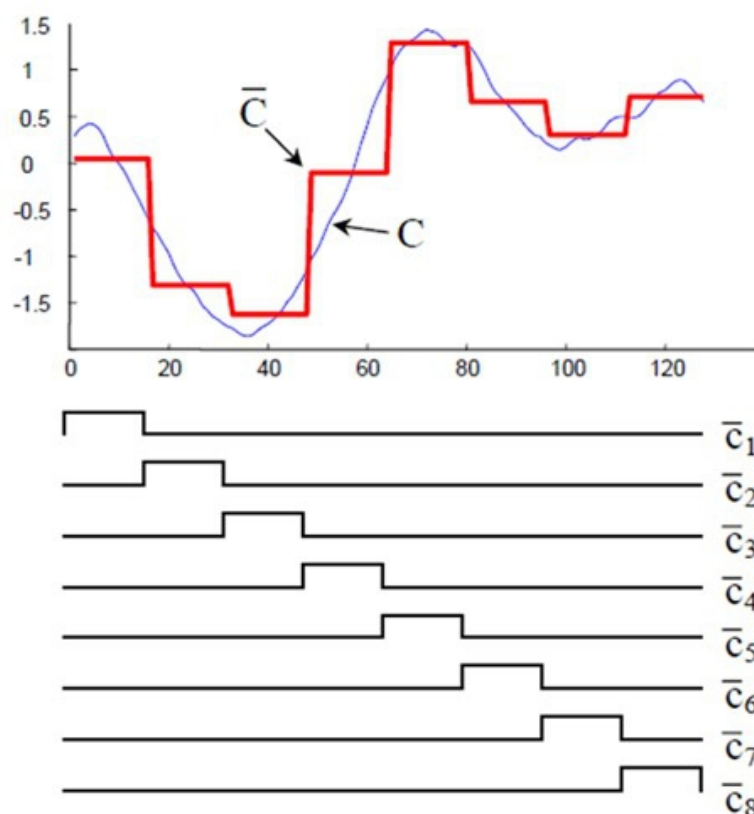
Em que,

$$j: \frac{n'}{n} * (t - 1) + 1$$

$$i: \frac{n'}{n} * t$$

Esta equação simplesmente indica que uma ST é dividida em n' janelas e que o valor médio representa todos os pontos dentro da janela. O parâmetro n' pode ser encontrado fazendo um gráfico do erro de reconstrução em função de N . Dessa forma, realiza-se a transformação de uma série temporal em uma série temporal na representação PAA, exibida por \bar{C} . A Figura 16 mostra essa transformação.

Figura 16 – Série temporal na representação PAA.



Fonte: (LIN et al., 2007)

Pode-se observar na Figura 16, que nesta representação transformada, têm-se uma redução dimensional de 128 dados para 8. Como mencionado anteriormente, a manipulação computacional da série temporal representada por \bar{C} , é visivelmente superior ao que seria aplicada diretamente na série temporal C . Por exemplo, para computar a relação entre duas curvas de ST, a redução do comprimento da série temporal n para n' pode levar a redução de $O(n)$ para $O(n')$ em termos de complexidade computacional. Sendo $n' \ll n$, a complexidade computacional será amplamente reduzida.

3.1.3.2 Adaptive Piecewise Aggregate Approximation

A representação *Adaptive Piecewise Aggregate Approximation* (APAA) é uma extensão da representação PAA (KEOGH et al., 2001). Contudo, ao invés de utilizar segmentos de mesmo tamanho, a APAA permite que a série temporal seja representada por segmentos ou janelas

de tamanhos distintos, ou seja a redução da dimensionalidade é realizada de acordo com o comportamento da série temporal em análise (FERREIRA et al., 2018).

A média de cada janela é baseada em técnicas de Change Point Detection (CPD) e no algoritmo PAA; o CPD consistem em variações abruptas nos dados da série temporal. Tais mudanças abruptas podem representar transições que ocorrem entre estados. A detecção desses pontos de mudança é útil na modelagem e previsão de séries temporais sendo bastante utilizada na detecção de alterações climáticas (AMINIKHANGHAHI e COOK, 2017). O objetivo do APAA é adaptar automaticamente o comprimento de cada janela à sua condição local de variabilidade e diferença, para o valor médio dos valores atuais em que cada janela é definida (FERREIRA et al., 2018).

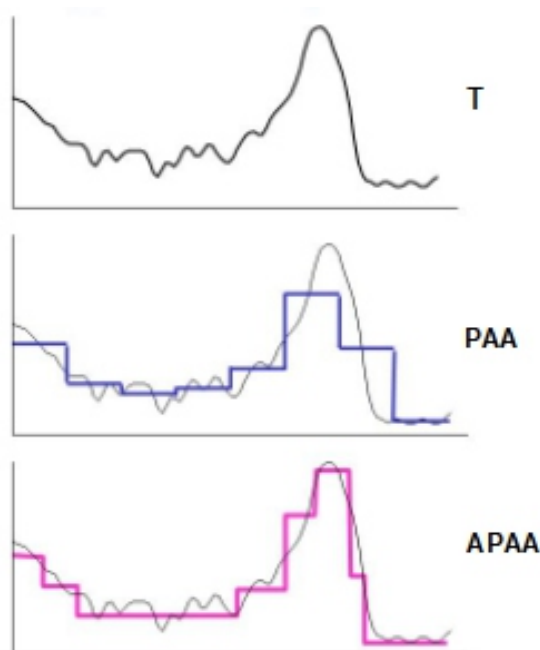
Diminuindo assim a possibilidade de que eventos que ocorram em intervalos de tempos distintos sejam desconsiderados devido ao tamanho fixo dos segmentos, como também que grande parte de uma série temporal, com baixa atividade, seja representada por um único segmento e que partes que possuem maior atividade e que sejam de maior interesse sejam representadas por um número maior de segmentos.

Desse modo, a redução de dimensionalidade por meio dessa técnica utiliza a localização dos picos das séries temporais para determinar o comprimento desses segmentos. Os parâmetros relacionados são a mínima distância entre os picos - minpeakdist e o tamanho mínimo do pico - minpeakh para ser considerado como um pico apropriado. Sendo esses parâmetros relacionados de acordo com os critérios relativos ao comprimento mínimo do segmento e à sensibilidade as variações existentes nas séries temporais. Após determinado o tamanho de cada janela é calculado o valor médio desses valores que irá representar o valor do patamar considerado. Na Figura 17 é ilustrada uma comparação da transformação de uma série temporal T , aplicando as técnicas PAA e APAA, para efeitos de comparação visual da eficiência de cada uma das técnicas, onde pode-se perceber que o tamanho de cada janela muda a partir do comportamento da ST.

Esse processo será repetido sucessivamente a partir do próximo valor usando uma nova janela vazia, até que todas os segmentos da ST sejam definidos. Desse processo de aproximação será obtido uma sequência de segmentos de tamanhos distintos, cada um contendo um número diferente de pontos de dados. E cada um desses segmentos será representado como descrito anteriormente.

Essa nova técnica alternativa quando comparada com outras técnicas existentes na literatura conclui-se que não apenas pode-se reduzir notoriamente a dimensionalidade da série temporal, como também permite que cada dimensão da nova ST possa representar eventos importantes presentes nessa série. Além disso, a técnica permite armazenar informações locais presentes nas ST, as quais poderiam ser combinadas com representações de comportamentos globais da série temporal com o intuito de obter uma análise mais completa.

Figura 17 – Comparação visual das técnicas de PAA e APAA para reduzir a dimensionalidade de séries temporais.



Fonte: Adaptado de KEOGH et al. (2001)

3.1.4 Regressão Quantílica

Uma das técnicas mais utilizadas no meio acadêmico é a Regressão Linear (RL), que tem como objetivo principal precisar uma relação entre a variável resposta e a variável preditora. A teoria clássica dos modelos lineares de regressão é essencialmente inspirado na teoria clássica das esperanças condicionais. A curva de regressão mostra um grande resumo das médias das atribuições correspondentes ao conjunto dos x 's observados; fazendo uma análise mais profunda é possível calcular diversas curvas de regressões diferentes, correspondendo aos diferentes pontos percentuais da distribuição e assim obter uma visão mais completa desse conjunto (KOENKER; BASSET, 1978).

Assim, como a média fornece uma visão incompleta de uma única distribuição, a curva de regressão também fornecerá uma visão incompleta correspondente para um conjunto de distribuições. Ou seja, de imediato é possível questionar se a esperança condicional ou outra medida de tendência central condicional caracteriza de forma adequada a relação estatística entre a variável resposta e as demais variáveis predictoras do estudo.

No modelo de Regressão Linear Simples (RLS), a relação entre a variável preditora X e a variável resposta Y é dada pela equação de uma reta e é representada por:

$$Y_i = \beta_0 + \beta_1 X_i + e_i, \quad (21)$$

em que, o intercepto β_0 e o coeficiente angular β_1 são constantes ou parâmetros desconhecidas e e_i são erros aleatórios.

Uma maneira de se estimar os parâmetros da Equação ?? é através do método dos Mínimos Quadrados Ordinários (MQO), no qual se adotam como estimativas dos parâmetros desconhecidos os valores que minimizam a soma de quadrados dos erros, ou seja, das diferenças entre as distâncias verticais das observações y_i e a reta estimada (GUJARATI, 2011) e (CECON et al., 2012).

Contudo, o MQO pode gerar algumas estimativas errôneas quando existem muitas *outliers* nos dados ou quando a distribuição de probabilidade é muito assimétrica. Isso acontece, pois nesse método são atribuídos os mesmos pesos a cada observação durante o processo de estimação dos parâmetros (DRAPER; SMITH, 1998). Desta forma, é possível que a relação entre a variável resposta e a variável preditora não seja estabelecida de forma correta, fazendo com que sejam feitas interpretações erradas dos resultados.

O fato da regressão quantílica ser considerada robusta a *outliers* e outras violações de hipóteses está ligado ao processo de análise de regressão linear, onde valores que se afastam do padrão tendem a distorcer a média; dessa forma, a regressão linear simples é mais sensível a esses valores que se afastam da média e é afetada por eles, enquanto no modelo de regressão quantílica isso não acontece (HAO; NAIMAN, 2007).

Por esse motivo, muitas vezes se faz necessário a utilização de modelos mais robustos com o intuito de obter resultados mais confiáveis ou que contribuam com maiores informações à análise. Para tal, KOENKER e BASSET (1978) propuseram uma abordagem mais geral, a regressão quantílica (RQ), que se baseia no método dos erros absolutos ponderados. Entretanto, nesta metodologia não se considera apenas o valor mediano, e sim é realizada uma ponderação na minimização dos erros para se estimar os diversos quantis de interesse.

A utilização de diversos quantis possibilita a obtenção de maiores informações de localização do que quando se utiliza apenas o centro da distribuição. Assim, a regressão quantílica pode ser caracterizada como um modelo de regressão semiparamétrico, pois não exige nenhuma distribuição de probabilidade para a variável resposta. BUCHINSKY (1998) sintetiza de forma objetiva algumas características importantes da regressão quantílica:

- O modelo pode ser utilizado para caracterizar toda distribuição condicional da variável resposta, dado um vetor de covariáveis independentes.
- Os modelos de RQ têm uma representação de programação linear que torna as estimativas mais fáceis.
- O estimador soma dos desvios ponderados é responsável por dar um caráter mais robusto à RQ, de modo que a estimativa do vetor de coeficientes não seja sensível a *outliers* na variável resposta.
- Soluções diferentes nos diversos quantis de interesse podem ser obtidas a partir da modificação das covariáveis nos diferentes pontos de distribuição condicional da variável resposta.

Seja Y uma variável aleatória com uma função distribuição acumulada $F(y) = \Pr(Y \leq y)$; então o quantil de ordem τ da variável Y pode ser definido como:

$$Q_\tau(Y) = F^{-1}(\tau) = \inf\{y | F(y) \leq \tau\}, \quad (22)$$

em que, $\tau \in [0, 1]$.

Por exemplo, para $\tau = 0.5$ tem-se o caso em que é estimado a média condicional (KOENKER; BASSET, 1982). A mediana é um quantil especial que descreve a localização central da distribuição. Portanto, a mediana condicional é um caso especial de regressão quantílica, onde o quinto quantil é modelado como uma função das covariáveis. Outros quantis também podem ser utilizados para descrever posições não centrais de uma distribuição.

Considere (y_t, x_t) , $t = 1, \dots, n$, uma amostra obtida de um conjunto de dados qualquer. É possível relacionar a variável resposta, y , e o vetor de covariáveis, x , por meio da seguinte relação linear:

$$Q_\tau(y_t | x_t) = \beta_{0(\tau)} + \beta_{1(\tau)}x_{t1} + \dots + \beta_{p(\tau)}x_{tp}, \quad (23)$$

Em que, $Q_\tau(y_t | x_t)$ é o quantil condicional de $y_t | x_t$ e $\beta_{0(\tau)}, \dots, \beta_{p(\tau)}$ são os parâmetros desconhecidos indexados no quantil τ . O quantil de ordem τ de uma amostra é um ponto tal que $100\tau\%$ dos valores amostrais são inferiores a ele, com $0 < \tau < 1$.

Esses parâmetros desconhecidos, indexados no quantil de ordem τ , podem ser estimados diretamente por meio da solução do problema de minimização definido na Equação 24 descrita em KOENKER e MACHADO (1999).

$$\frac{\min}{(\beta \in \mathcal{R}^p)} \sum_{t=1}^n \rho(\tau)(\beta_{0(\tau)} + \beta_{1(\tau)}x_{t1} + \dots + \beta_{p(\tau)}x_{tp}), \quad (24)$$

Em que, $\rho(\tau)$ é um termo ponderador. Desta forma o τ -ésimo estimador $\beta'(\tau)$ da regressão quantílica é escolhido de forma a minimizar a soma dos desvios do y observado até um ajuste y' (NEYMAN; PEARSON, 1928). Assim, pode-se reescrever o modelo da seguinte maneira:

$$\sum_{t=1}^n d(y_t, y'_t) = \tau \left(\sum_{y_t \geq y'_t} |y_t - x_t^T \beta'_\tau| \right) + (1 - \tau) \left(\sum_{y_t < y'_t} |y_t - x_t^T \beta'_\tau| \right), \quad (25)$$

Em que, $d(y_t, y'_t)$ é a distância que deve ser minimizada.

Generalizando esse resultado para qualquer quantil de interesse, $\tau \in [0, 1]$, tem-se que:

$$d_\tau(y_t, y'_t) = \begin{cases} (1 - \tau)|y_t - x_t^T \beta'_\tau| & \text{se } y_t < y'_t \\ \tau|y_t - x_t^T \beta'_\tau| & \text{se } y_t \geq y'_t \end{cases} \quad (26)$$

Em que, função de distância multiplica os resíduos por τ se eles forem não-negativos e por $\tau-1$ caso contrário, para que desta forma sejam tratados assimetricamente. Ou seja, o método de estimação da regressão quantílica busca atribuir pesos diferentes a um determinado conjunto de observações dependendo de sua localidade na reta e regressão estimada. Assim, a estimação

para cada regressão é baseada nos dados ponderados de toda a amostra em vez de considerar uma porção amostral daquele quantil (NEYMAN; PEARSON, 1928).

Segundo KOENKER e BASSET (1978), as principais vantagens desse método sobre a técnica dos mínimos quadrados são:

- A técnica de regressão quantílica permite caracterizar toda a distribuição condicional de uma variável de resposta a partir de um conjunto de regressores;
- Regressão quantílica pode ser utilizada mesmo quando a distribuição não é Gaussiana;
- Regressão quantílica utiliza a totalidade dos dados para estimar os coeficientes angulares dos quantis, ou seja, não há sub amostras dos conjuntos de dados;
- A regressão quantílica é robusta a *outliers*;
- Por utilizar a distribuição condicional da variável resposta, podem ser estimados os intervalos de confiança dos parâmetros e da regressão diretamente dos quantis condicionais desejados;
- Como os erros não possuem uma distribuição normal, os estimadores provenientes da regressão quantílica podem ser mais eficientes que os estimadores por meio do método dos mínimos quadrados;
- A regressão quantílica pode ser representada como um modelo de programação linear, o que facilita a estimação dos parâmetros, e diversos softwares econométricos possuem ferramentas para este tipo de análise.

3.1.4.1 Modelo de Regressão Quantílica

KOENKER e BASSET (1978) iniciaram os estudos sobre a regressão quantílica, que modela quantis condicionais como funções de preditores. O modelo de regressão quantílica é uma extensão natural do modelo de regressão linear. Enquanto a regressão linear especifica mudanças na média das variáveis dependentes, associada a mudanças nas covariáveis, o modelo de regressão quantílica especifica mudanças em um determinado quantil. Considerando que qualquer quantil possa ser utilizado, o modelo permite pré-determinar qualquer posição na distribuição. A regressão quantílica permite aos pesquisadores escolher posições na amostra que atendam aos seus objetivos (HAO; NAIMAN, 2007).

O modelo da Regressão Quantílica, que descreve a relação funcional entre a variável dependente e a variável independente, pode ser descrito como:

$$Y'_t = \beta_{0i}(\tau) + \beta_{1i}(\tau)Y_t + e_i(\tau), \quad (27)$$

em que,

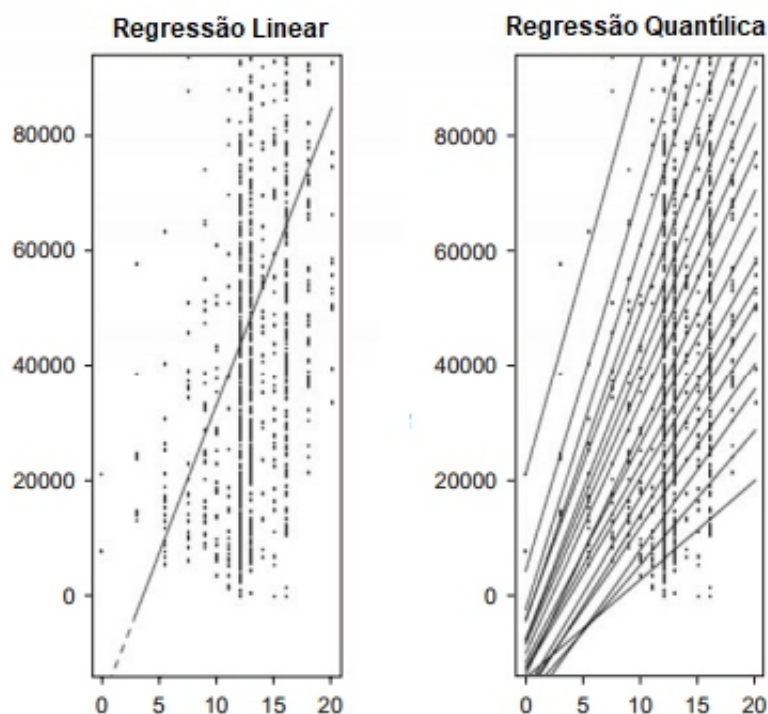
- $\beta_{0i}(\tau)$: é a constante da regressão;

- $\beta_{li}(\tau)$: é o coeficiente de regressão;
- $e_i(\tau)$: são os erros aleatórios independentes e identicamente distribuídos com quantil de ordem τ igual a zero;
- Y_i : é variável independente.

Para os modelos de RLS é estimada apenas uma reta para explicar todo o conjunto de dados. Contudo, a linha da regressão não irá identificar alterações na distribuição da variável independente. Já no modelo de RQ, são estimadas retas para cada quantil de interesse; dessa forma, a interpretação se tornará mais adequada à interpretação dos resultados para o conjunto de dados com a presença de assimetria, pois através dela é possível traçar a relação em regiões centrais, através da mediana, e nas caudas da distribuição condicional, de acordo com o interesse.

Na Figura 18, pode-se observar que a RLS propicia apenas a visualização em termos médios da variável dependente, enquanto que a RQ fornece informações sobre a relação funcional em diferentes níveis da variável dependente.

Figura 18 – Ajuste de um Modelo Linear e diversos ajustes da Regressão Quantílica



Fonte: (HAO; NAIMAN, 2007)

4 RESULTADOS E DISCUSSÕES

Este trabalho propõe uma comparação entre três métodos de análise de grandes bancos de dados de séries temporais de velocidade do vento, com o objetivo de identificar os valores extremos de velocidade do vento nas bases de dados analisadas; esses valores extremos correspondem aos valores iguais ou superiores ao *cut-out* (20 m/s) e iguais ou inferiores ao *cut-in* (3 m/s) da turbina eólica.

Para tal, utilizou-se os métodos de redução de dimensionalidade de ST, PAA e APAA, visando menor esforço computacional e maior rapidez na análise dos banco de dados. Também a Regressão Quantílica, em conjunto com o método PAA, em que permitiu-se a realização de uma ponderação na minimização dos erros para se estimar os diversos quantis de interesse, que no nesse caso foram os quantis 0,05 e 0,95 que representaram os valores extremos de mínimo e máximo respectivamente.

Para o método PAA, os dados foram agrupados em janelas com 4 medições de 10 minutos cada, gerando novos dados em diferentes escalas de uma fonte comum. Com relação ao método PAA em conjunto com a regressão quantílica, para cada janela obtida a partir do método PAA, foram extraídos os 5% maiores valores e 5% menores valores desse intervalo, que no caso correspondem ao menor e maior valor da janela, já que tem-se apenas quatro valores em cada janela. Já o método APAA permite que as séries temporais sejam representadas por janelas de tamanho variável, as quais variam de acordo com as características da ST.

Dessa forma, é possível por meio dessa análise identificar se determinada região é favorável ou não à implantação de uma parque eólico, como também determinar os períodos do ano em que a turbina eólica ficará fora de operação, e assim pode-se realizar o planejamento e programação da operação energética, principalmente nos subsistemas sul e nordeste, que são as regiões onde estão localizados os parques eólicos onde foram coletados os dados analisadas. A implementação computacional foi feita por meio da ferramenta MATLAB[®], aonde foram gerados códigos para cada um dos três métodos abordados.

4.1 Descrição dos Dados

O estudo foi realizado utilizando dados de duas regiões, o Sítio Eólico da Região Sul (SERS) e o Sítio Eólico da Região Nordeste (SERNE). As bases utilizadas já levam em consideração a medida da velocidade do vento na altura do cubo da turbina eólica, ou seja, à altura de 100 m. Os sítios foram escolhidos para representar melhor uma variedade de cenários ao estudar problemas de gerenciamento de velocidade do vento no Brasil, pois os dois possuem características bem divergentes. As séries temporais utilizadas neste trabalho consistem em valores de velocidade do vento medidos a cada 10 minutos, começando às 0h de 1 de janeiro de 2014 e terminando às 23h de 31 de dezembro de 2016. Os bancos de dados correspondem a

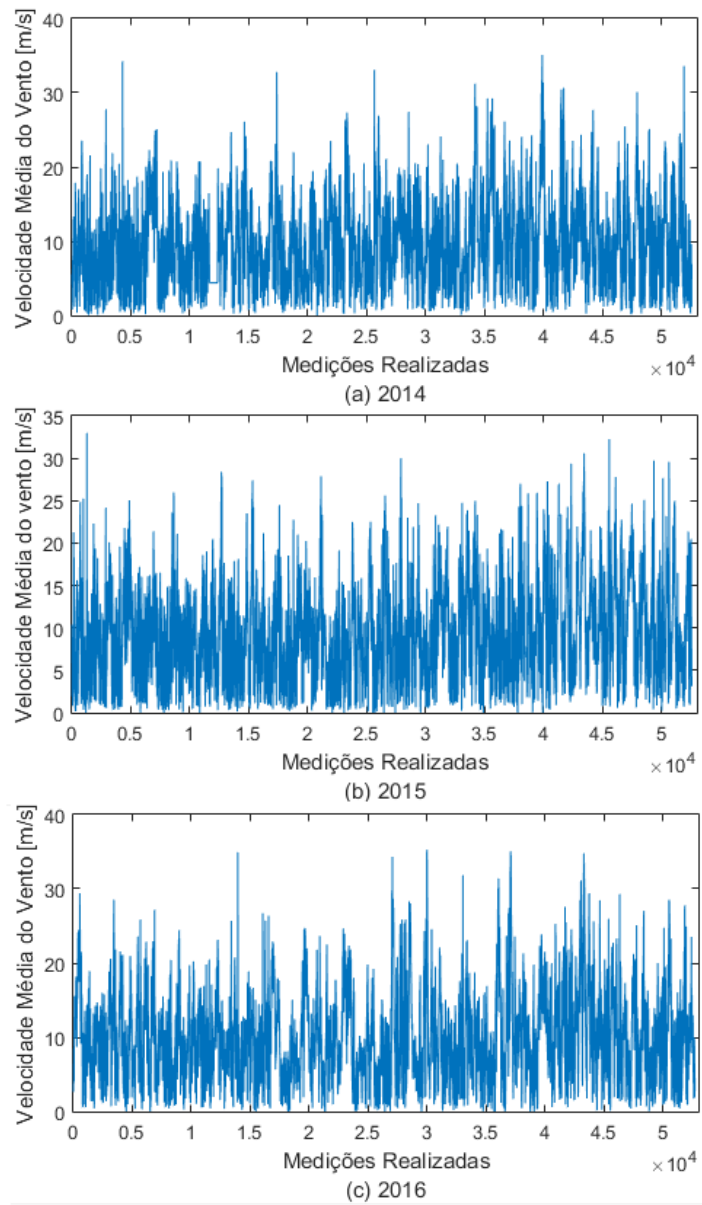
medidas obtidas do Operador Nacional do Sistema Elétrico (ONS) das regiões Sul e Nordeste.

Existem dois tipos principais de eventos extremos ao gerar energia eólica que são conhecidos como ocorrências de *cut-out* e *cut-in*. As ocorrências de *cut-out* são velocidades de vento maiores ou iguais a 20 m/s, as quais nomeou-se no decorrer do trabalho como velocidades de UP. Quando a velocidade do vento é igual ou superior a esse valor, as turbinas devem ser desligadas por razões de segurança estrutural das mesmas. Já as ocorrências de *cut-in* são velocidades do vento inferiores a 3 m/s, as quais nomeou-se no decorrer do trabalho como velocidades de LOW. Neste caso, as turbinas não estão em operação porque suas lâminas não têm velocidade de partida suficiente.

Uma verificação de qualidade indicou alguns valores inexistentes, os quais foram denominados como NaN (*Not a Number*). Eles correspondem a valores de medição perdidos que podem comprometer a sazonalidade da série de tempo de velocidade do vento. Como uma alternativa para resolver este problema, os NaN são ajustados para valores correspondentes às médias horárias referentes a cada uma das medições realizadas a cada 10 minutos para cada hora de cada dia do mês em estudo. Caso não exista nenhuma valor para a medição realizada a cada 10 minutos, é utilizada a média horária do ano anterior ao ano em análise.

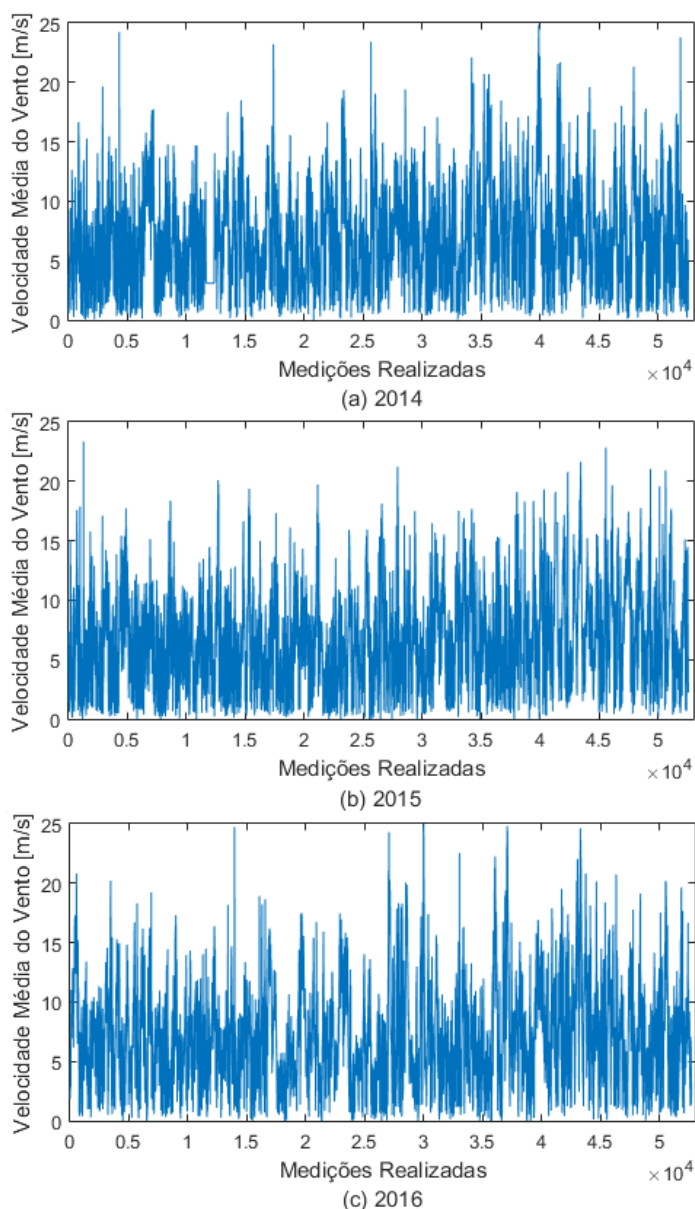
As Figuras 19 e 20 mostram o comportamento das séries temporais de vento dos Sítios Eólicos analisados nos três anos em estudo, 2014, 2015 e 2016, após a substituição dos NaN pelos valores das médias horárias para cada uma das medições realizadas a cada 10 minutos, buscando reduzir possíveis distorções na sazonalidade nos valores de velocidade.

Figura 19 – Série temporal para o SERS referente aos anos 2014, 2015 e 2016.



Fonte: Próprio Autor

Figura 20 – Série temporal para o SERNE referente aos anos 2014, 2015 e 2016.



Fonte: Próprio Autor

4.2 Imputação de Valores Ausentes

Um problema significativo e frequentemente encontrado em bancos de dados de velocidade do vento (e especificamente nos parques eólicos do Brasil), é a ausência de medidas de vento com qualidade suficiente para uso em projetos elétricos de energia eólica. Assim, o primeiro passo de pré-processamento dos dados é a identificação dos valores em falta nas bases de dados. Isso é feito através do desenvolvimento de algoritmos que detectem a ausência de valores nas bases de dados. Posteriormente é feito o cálculo dos valores das médias horárias de cada uma das medições realizadas em 10, 20, 30, 40, 50 e 60 minutos em cada uma das 24 horas do dia de cada mês. Assim, é necessário que a base tenha ao menos um valor referente a cada

uma dessas medições para que a média seja calculada, caso contrário será utilizada os valores das medições do respectivo mês do ano anterior. E por fim as lacunas que foram identificadas nas bases de dados são substituídas pelos valores das médias calculadas. Isso para obter um conjunto consistente de dados para que os resultados estejam mais próximos da realidade, isto é, preservando sua sazonalidade.

4.3 Padrões de Velocidades Analisados

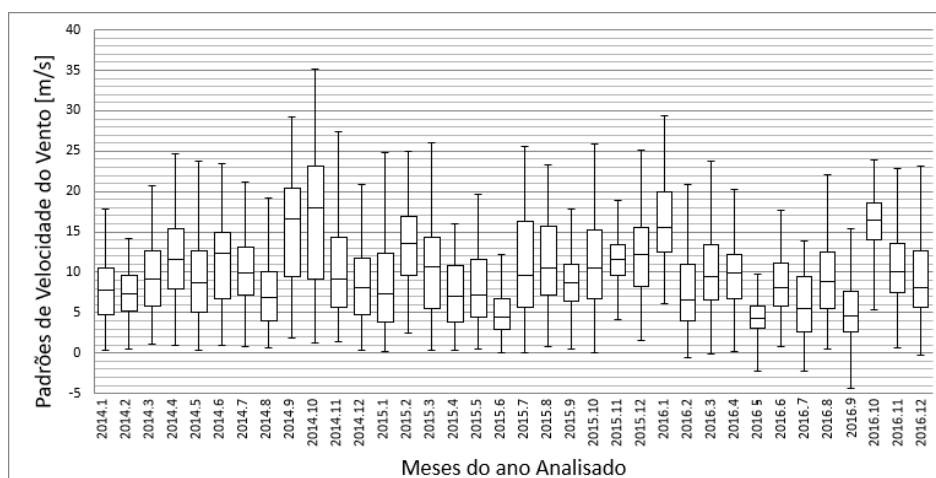
Depois de preencher os valores faltantes da série com os valores das médias diárias das velocidades do vento, foram construídos *boxplots* para verificar se os padrões de velocidade do vento estão dentro das velocidades do estudo, que são velocidades de LOW e UP consideradas.

Para o SERS, observa-se que, nos valores analisados, há padrões de interesse abaixo de 3 m/s e acima de 20 m/s. Para o ano de 2014, os meses de setembro, outubro e novembro apresentam a maioria dos valores acima de 20 m/s, o mês de outubro mostrou padrões de velocidade do vento acima de 30 m/s. Os meses de janeiro, fevereiro, maio e dezembro apresentam padrões de vento com valores bem abaixo de 3 m/s, perto de 0 m/s. No ano de 2015, pode-se verificar que os padrões de vento acima de 25 m/s ocorrem nos meses de março, julho e outubro, mas não atingem 30 m/s como em 2014; já os meses de junho e julho apresentam os valores mais baixos para a velocidade do vento. Em 2016, vários meses têm padrões de vento acima de 20 m/s, mas apenas o mês de janeiro apresentou valores de velocidade do vento acima de 25 m/s. Os padrões mais baixos de velocidade do vento ocorrem nos meses de maio, julho e setembro. Como pode ser observado na Figura 21.

Para o SERNE, observou-se que a maioria dos padrões de vento de interesse da análise são valores inferiores a 3 m/s, uma vez que em nenhum dos anos analisados foram observados valores superior a 20 m/s. No ano de 2014, os meses de março, maio e setembro mostraram um maior número de padrões abaixo de 3 m/s; já em 2015, esses valores foram observados nos meses de fevereiro, março e maio. O ano de 2016 apresentou padrões de vento abaixo de 3 m/s nos meses de fevereiro, março e junho. Como pode ser observado na Figura 22.

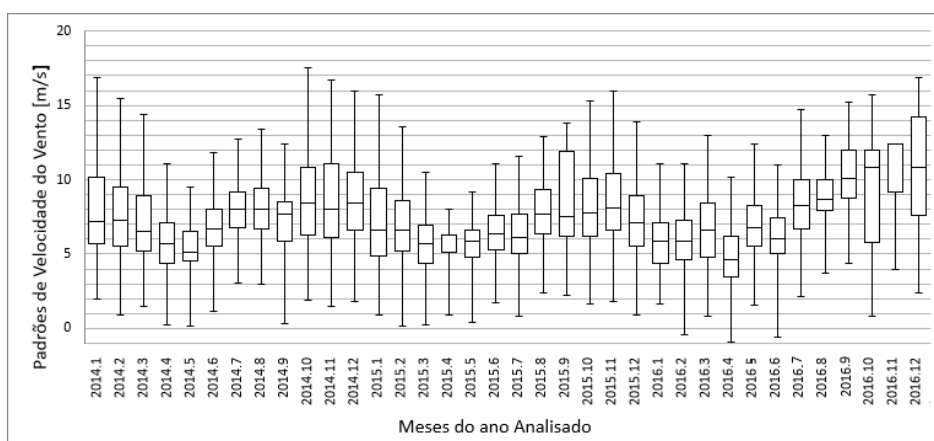
Também pode ser observado que a maioria dos padrões nos três anos em estudo estão dentro do alcance operacional seguro para as turbinas eólicas nos dois sítios analisados.

Figura 21 – BoxPlot - Padrões de velocidade do vento para os anos 2014, 2015 e 2016 para o SERS



Fonte: Próprio Autor

Figura 22 – BoxPlot - Padrões de velocidade do vento para os anos 2014, 2015 e 2016 para o SERNE



Fonte: Próprio Autor

4.4 Aplicação dos Métodos Propostos

Este trabalho propõe uma representação PAA das séries temporais de velocidades, a qual em um dia tem-se 144 medições (medições feitas a cada 10 minutos em um dia) com a decomposição da série que corresponde a um dia em 36 intervalos de tempo (janelas) com 4 medições realizadas a cada 10 minutos, gerando novos dados em diferentes escalas de uma fonte comum. E uma representação APAA assemelha-se à representação PAA, mas em vez de usar segmentos do mesmo tamanho, a APAA permite que as séries temporais sejam representadas por janelas de tamanho variável, portanto regiões menos detalhadas podem ser representadas por uma única janela e regiões com maiores detalhes usam um maior número de janelas. A proposta analisa as séries temporais de velocidade do vento para modelar os recursos associados aos seus valores extremos.

Em primeiro lugar, os dados obtidos com a aplicação do algoritmo PAA são analisados. A abordagem PAA é então aprimorada e a RQ é utilizada sobre a partição PAA para lidar com os quantis superiores e inferiores para cada janela. Nesta análise, considera-se 0,95 como quantile superior e 0,05 como quantile inferior. Assim, existe 5% dos valores mais altos de velocidade do vento sobre quantile 0,95 e 5% dos valores inferiores abaixo de quantile 0,05. No presente estudo existem 4 medições realizadas a cada 10 minutos em uma janela, os valores escolhidos para os quantiles superiores e inferiores são os valores mais altos e mais baixos do PAA de cada janela. A Tabela 3 mostra os tamanhos de janelas usados, m representa o número de amostras e n o total de dias de cada mês dos anos analisados.

Tabela 3 – Quantidade de janelas para cada conjunto de dados

Quantidade de dias do mês	n	m
31	4464	1116
30	4320	1080
29	4176	1044
28	4032	1008

Fonte: Próprio Autor

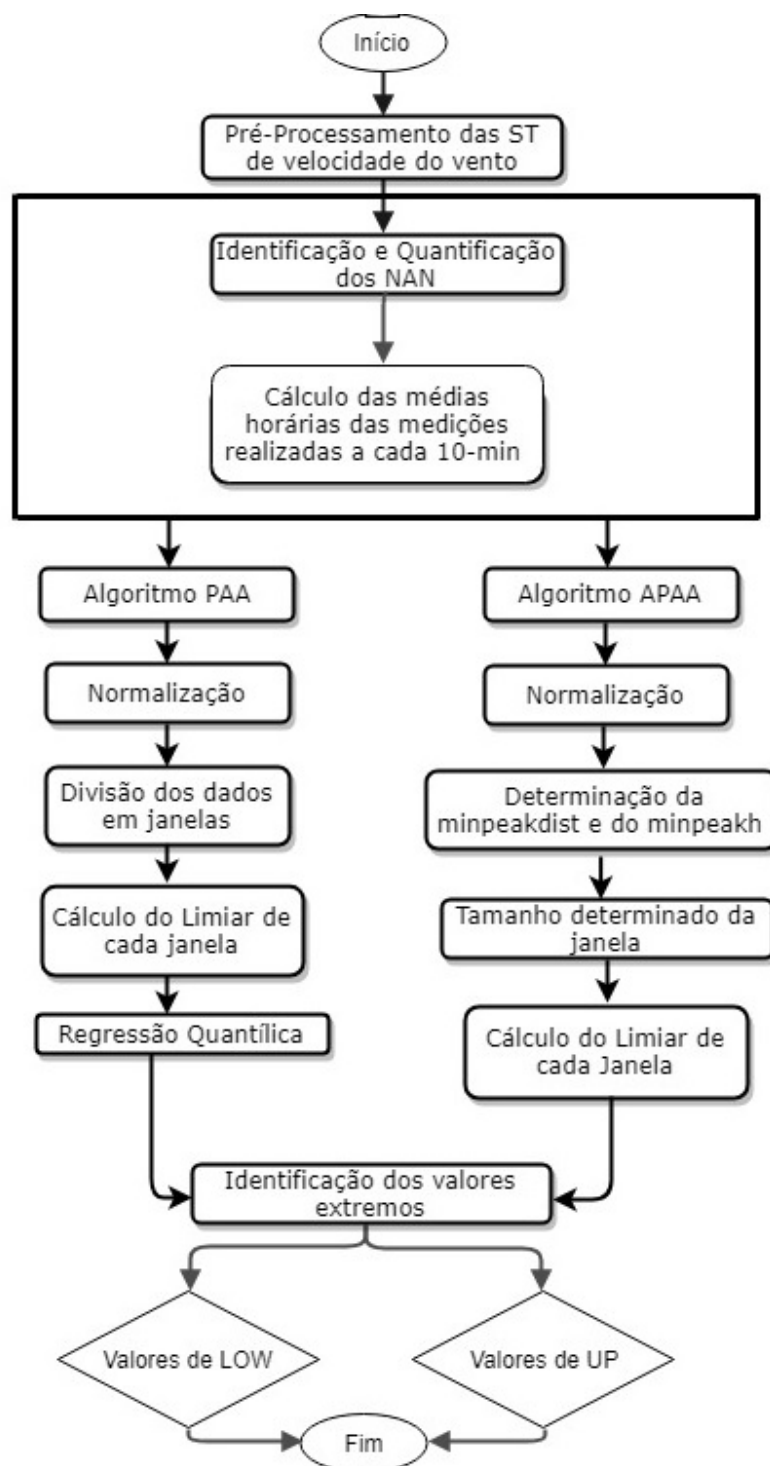
Posteriormente, foi proposto um método alternativo ao PAA e PAA/RQ, o algoritmo APAA, onde a redução da dimensionalidade varia com o comportamento das séries temporais, não sendo necessário a determinação do tamanho da janela como no PAA; seu tamanho varia a partir da relevância da informação, isso permite que os eventos que ocorrem em diferentes intervalos de tempo do segmento dado sejam desconsiderados. Pode observar-se nas Tabelas 4 à 7, a quantidade de valores extremos UP e LOW para cada uma das regiões analisadas após a implementação dos métodos propostos. Os valores empíricos correspondem às quantidades reais identificadas em cada um dos parques eólicos. A Figura 23 apresenta um fluxograma, resumindo as técnicas aplicadas na análise das séries temporais de velocidade do vento dos sites estudados. Ela abrange uma série de etapas para obter os resultados de interesse.

A aplicação dos métodos utilizados pode ser facilmente visualizada nas Figuras 24, 25 e 26, como uma modelagem de uma sequência com uma combinação linear de funções de base de janela. Nessas figuras tem-se a representação de 300 medições do mês de janeiro referente ao ano de 2014, onde são mostradas as ocorrências de eventos de LOW e UP para os três métodos implementados.

Para o algoritmo PAA e PAA/RQ, os registros dessas ocorrências são feitos com janelas de 4 medidas de 10 min cada, onde se tem 96 registros diários e sua quantificação é feita respeitando sempre os intervalos das janelas. Para o algoritmo APAA, são utilizados 2 valores para cada janela, um representou a média e o outro, o tamanho que nos permitiu estimar o comprimento de cada janela. Para o método PAA, pode-se notar que não houve a detecção de ocorrências no intervalo determinado; diferentemente do método PAA, o PAA/RQ e APAA apresentaram várias

ocorrências. As janelas são representadas pela cor verde e e, quando a mudança de cor ocorre, o método em questão identifica um registro de evento extremo nos sítios eólicos analisados.

Figura 23 – Fluxograma com as etapas de cada método aplicado



Fonte: Próprio Autor

Tabela 4 – Demonstrativo dos valores obtidos de ocorrências LOW e UP para as técnicas PAA/RQ e APAA para o SERS

Sítio Eólico da Região Sul (SERS)													
Mét.	Empírico						PAA						
	LOW			UP			LOW			UP			
Occ.													
Ano	1	2	3	1	2	3	1	2	3	1	2	3	
Jan	502	784	309	64	65	254	118	184	77	16	15	58	
Fev	393	784	553	123	50	103	85	133	128	49	30	26	
Mar	368	545	318	7	122	79	78	125	74	0	29	18	
Abr	375	540	385	95	100	87	92	151	93	22	24	22	
Mai	557	652	425	36	75	119	134	119	107	9	17	29	
Jun	417	518	517	196	61	163	93	192	129	46	15	45	
Jul	229	797	513	38	189	581	49	166	128	5	44	147	
Ago	466	697	489	223	137	64	110	76	113	55	28	15	
Set	218	310	583	351	136	423	58	76	149	85	32	106	
Out	239	318	132	591	661	583	65	38	33	153	164	142	
Nov	321	187	384	196	241	147	72	49	101	47	56	33	
Dez	443	213	396	196	170	219	96	79	91	45	45	54	
Total	4528	5890	5004	2109	2007	2822	1050	1388	695	532	499	1223	

1: ano 2014; 2: ano 2015; 3: ano 2016

Fonte: Próprio Autor

Tabela 5 – Demonstrativo dos valores obtidos de ocorrências LOW e UP para as técnicas PAA/RQ e APAA para o SERS

Sítio Eólico da Região Sul (SERS)													
Mét.	PAA/RQ						APAA						
	LOW			UP			LOW			UP			
Occ.													
Ano	1	2	3	1	2	3	1	2	3	1	2	3	
Jan	186	302	115	27	27	87	415	765	369	54	90	385	
Fev	193	227	189	56	40	40	282	670	471	122	0	72	
Mar	140	198	114	5	44	31	350	772	295	0	149	99	
Abr	128	263	158	43	41	36	430	717	466	139	176	62	
Mai	211	198	152	11	27	45	338	682	401	96	140	302	
Jun	164	275	193	80	25	65	316	945	689	246	0	106	
Jul	99	232	169	19	60	164	152	980	598	0	218	573	
Ago	166	121	166	65	57	27	551	135	549	230	0	44	
Set	99	126	200	137	50	126	226	149	903	370	226	498	
Out	91	085	50	180	194	192	250	167	81	691	720	828	
Nov	131	081	64	83	97	188	183	167	351	257	0	164	
Dez	183	125	163	83	57	70	367	264	243	226	0	316	
Total	1791	2233	1733	789	719	1071	3860	6413	5416	2431	1719	5416	

1: ano 2014; 2: ano 2015; 3: ano 2016

Fonte: Próprio Autor

Tabela 6 – Demonstrativo dos valores obtidos de ocorrências LOW e UP para as técnicas PAA/RQ e APAA para o SERNE

Sítio Eólico da Região Nordeste (SERNE)												
Mét.	Empírico						PAA					
Occ.	LOW			UP			LOW			UP		
Ano	1	2	3	1	2	3	1	2	3	1	2	3
Jan	39	20	344	0	0	0	3	3	66	0	0	0
Fev	50	307	70	0	0	0	9	64	14	0	0	0
Mar	51	344	152	0	0	0	3	66	35	0	0	0
Abr	196	191	145	0	0	0	31	35	27	0	0	0
Mai	290	14	53	0	0	0	60	2	9	0	0	0
Jun	60	81	35	0	0	0	8	15	5	0	0	0
Jul	40	46	8	0	0	0	6	8	0	0	0	0
Aug	4	4	0	0	0	0	0	0	0	0	0	0
Set	5	3	4	0	0	0	0	0	0	0	0	0
Out	4	1	4	0	0	0	1	0	1	0	0	0
Nov	19	12	0	0	0	0	2	0	1	0	0	0
Dez	11	31	5	0	0	0	2	3	0	0	0	0
Total	769	1054	820	0	0	0	125	196	158	0	0	0

1: ano 2014; 2: ano 2015; 3: ano 2016

Fonte: Próprio Autor

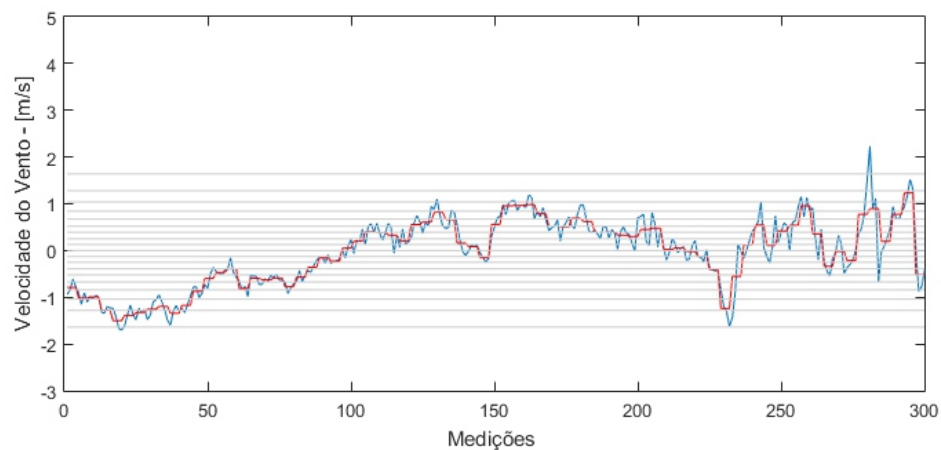
Tabela 7 – Demonstrativo dos valores obtidos de ocorrências LOW e UP para as técnicas PAA/RQ e APAA para o SERNE

Sítio Eólico da Região Nordeste (SERNE)												
Mét.	PAA/RQ						APAA					
	LOW			UP			LOW			UP		
	1	2	3	1	2	3	1	2	3	1	2	3
Jan	21	9	147	0	0	0	59	0	278	0	0	0
Fev	26	96	31	0	0	0	58	372	99	0	0	0
Mar	35	138	156	0	0	0	66	375	188	0	0	0
Abr	91	89	60	0	0	0	130	85	371	0	0	0
Mai	130	7	27	0	0	0	292	0	46	0	0	0
Jun	35	37	17	0	0	0	60	74	24	0	0	0
Jul	20	25	7	0	0	0	22	15	0	0	0	0
Ago	3	2	0	0	0	0	0	0	0	0	0	0
Set	4	4	2	0	0	0	0	0	0	0	0	0
Out	2	0	1	0	0	0	0	0	0	0	0	0
Nov	12	8	0	0	0	0	0	0	0	0	0	0
Dez	8	16	4	0	0	0	0	22	0	0	0	0
Total	387	431	452	0	0	0	687	943	1006	0	0	0

1: ano 2014; 2: ano 2015; 3: ano 2016

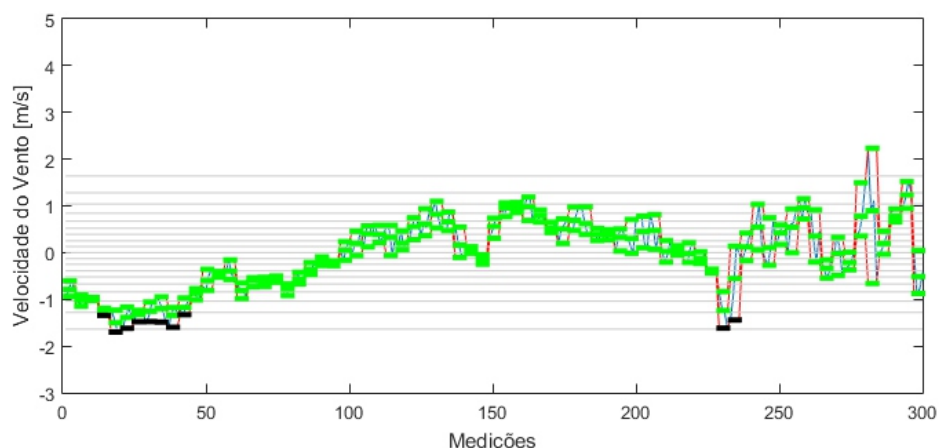
Fonte: Próprio Autor

Figura 24 – Representação dos resultados PAA para os valores extremos UP e LOW



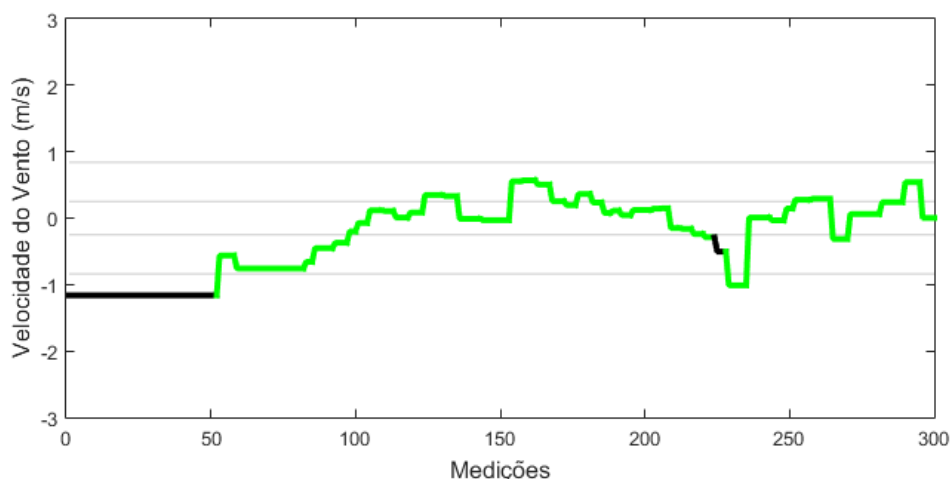
Fonte: Próprio Autor

Figura 25 – Representação dos resultados PAA/RQ para os valores extremos de UP e LOW



Fonte: Próprio Autor

Figura 26 – Representação dos resultados APAA para os valores extremos de UP e LOW



Fonte: Próprio Autor

4.5 Resultados

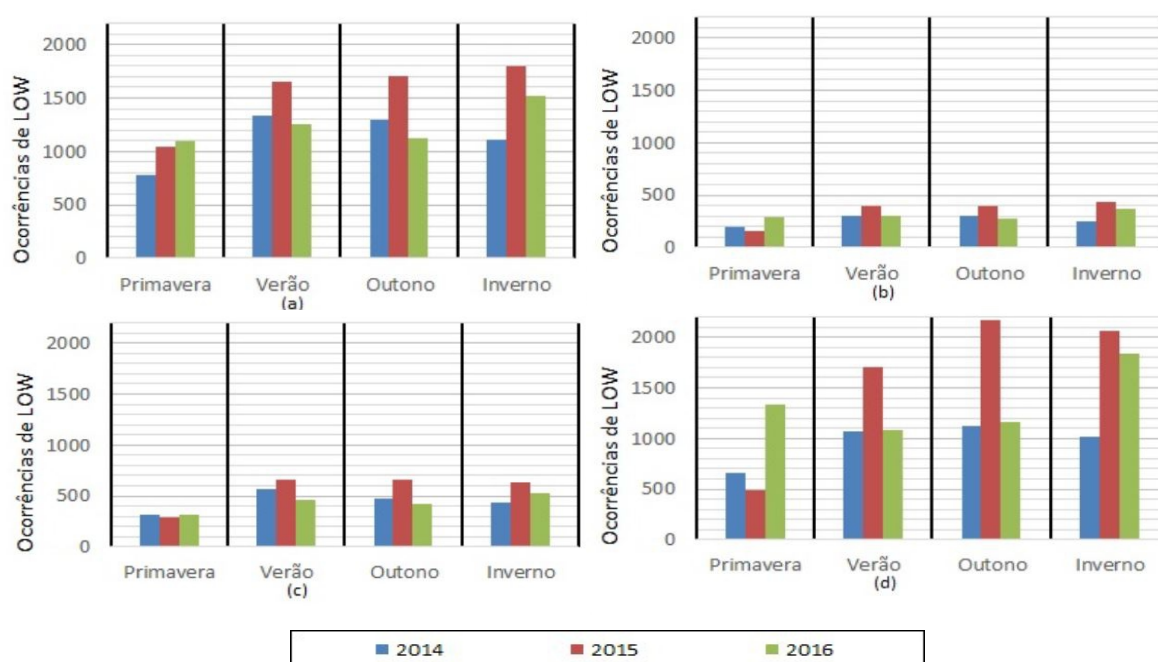
Esta seção apresenta os resultados da análise experimental realizada para avaliar o desempenho da metodologia proposta. Os resultados indicam que o APAA apresentou melhor desempenho do que PAA e PAA/RQ para os dois sítios eólicos. As análises são baseadas nas estações, devido à localização geográfica do Brasil; a primavera compreende os meses de setembro, outubro e novembro; o verão corresponde aos meses de dezembro, janeiro e fevereiro; o outono os meses de março, abril e maio, e finalmente o inverno com os meses de junho, julho e agosto. Assim, estima-se os períodos do ano de maior operação da turbina e os períodos em que se tem uma operação crítica.

Para as ocorrências de LOW do SERS, o ano de 2014 teve a maioria dos eventos extremos nas estações do verão, outono e inverno com mais de 82% das ocorrências, como se pode constatar na Figura 27 (a). Todos os métodos apresentaram a maior parte das ocorrências dos eventos extremos nas mesmas estações que os valores empíricos. Na técnica PAA o quantitativo dos valores extremos foi bem abaixo do esperado, totalizando apenas 23% da quantidade total esperada para o ano em estudo e da quantidade total das estações com maiores índices de valores extremos. A técnica PAA/RQ apresentou um maior número de ocorrências que a técnica PAA, correspondendo a quase 40% do total de ocorrências no ano em estudo e do total de ocorrências nas 3 estações citadas. Com relação a técnica APAA, o seu quantitativo chegou bem próximo do esperado, correspondendo a mais de 85% dos valores de ocorrências esperadas tanto no ano em estudo quanto nas estações com maior índice de eventos.

Já o ano de 2015 teve a maioria dos eventos extremos nas mesmas estações que o ano de 2014, ou seja, verão, outono e inverno com 88% das ocorrências, como se pode constatar na Figura 27 (a). Todos os métodos apresentaram a maior parte das ocorrências dos eventos extremos nas mesmas estações que os valores empíricos. Contudo, a técnica PAA apresentou menos de 24% do quantitativo total de ocorrências para o ano em estudo e para as estações com maior índice de ocorrências de valores extremos. Já para a técnica PAA/RQ o total de ocorrências foi superior ao PAA, todavia ainda manteve um índice baixo, com 40% do quantitativo total de ocorrências do ano em estudo como também para as estações que apresentaram os maiores índices de ocorrências de valores extremos. A técnica APAA identificou uma faixa de 8% a mais de ocorrências que os valores empíricos; apesar disso, ele ainda apresentou um desempenho mais satisfatório que as outras técnicas, chegando bem próximo do número de ocorrências dos valores empíricos.

No ano de 2016, os eventos extremos ocorreram de forma equilibrada nas 4 estações do ano como pode-se notar na Figura 27 (a). Todos os métodos também apresentaram uma distribuição uniforme do quantitativo de eventos extremos e cada uma das estações do ano em estudo. Contudo, a técnica PAA identificou um número bem abaixo de ocorrências quando comparado aos valores empíricos, correspondendo a menos de 25% do total. Já o método PAA/RQ teve um desempenho um pouco superior ao PAA apresentando quase 35% do total de ocorrências do ano em estudo. A técnica APAA apresentou, assim como no ano de 2015, uma faixa de 8% a mais de ocorrências que os valores empíricos, mesmo assim ele apresentou um desempenho mais satisfatório que as outras duas técnicas analisados, pois o quantitativo de ocorrências identificadas foi bem próximo dos valores empíricos.

Figura 27 – Gráficos de múltiplas barras de ocorrências LOW dos anos de 2014, 2015 e 2016 para SERS com (a) Valores empíricos, (b) Técnica PAA/RQ, (c) Técnica PAA e (d) Técnica APAA



Fonte: Próprio Autor

Para as ocorrências de UP do SERS, o ano de 2014 apresentou a maioria dos eventos extremos nas estações primavera, verão e inverno com quase 94% das ocorrências, com a primavera apresentando quase 60% dessas ocorrências, como é mostrado na Figura 28 (a). Com relação aos métodos analisados, todos eles identificaram as mesmas estações dos valores empíricos como estações com maior número de ocorrências. O método PAA identificou um quantitativo de ocorrências bem abaixo do esperado, com apenas 25% do total de ocorrências do ano em estudo e para as estações com maior índice de ocorrências de valores extremos.

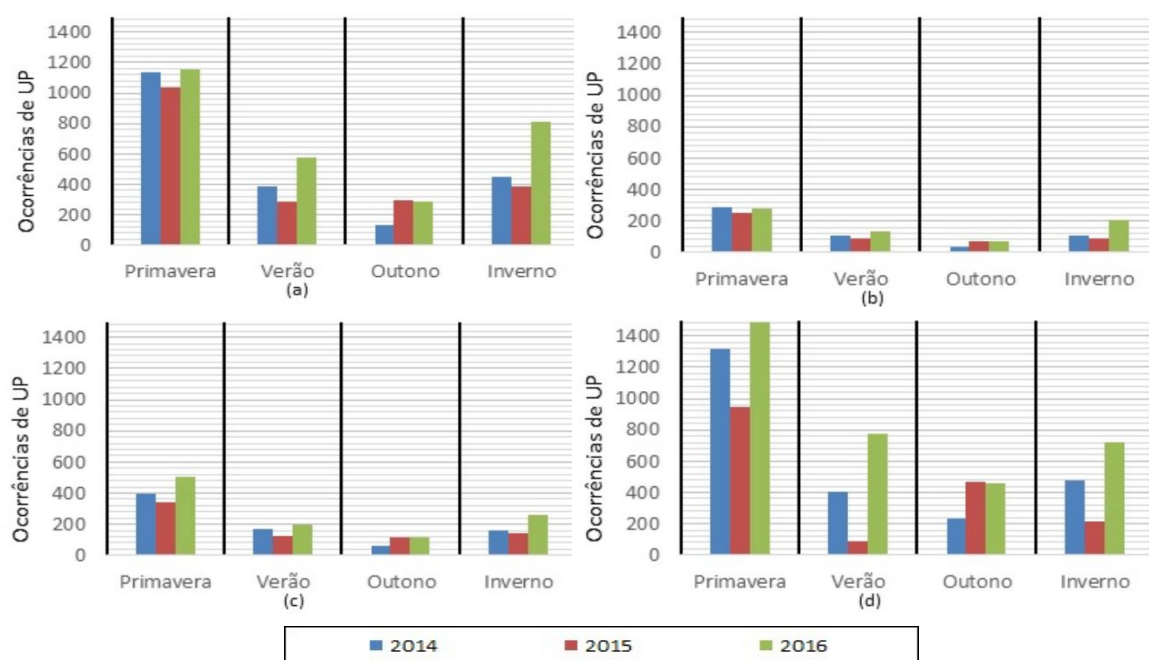
Já a técnica PAA/RQ identificou um número maior de ocorrências que o PAA, contudo ainda foi bem abaixo do esperado com apenas 37% da quantidade total de ocorrências identificadas nesse ano e quase 35% das ocorrências identificadas nas estações com maior quantitativo de valores extremos. Para a técnica APAA, os resultados foram mais satisfatórios que as outras técnicas pois a quantidade de valores identificados em cada uma das estações foi mais próxima dos valores esperados. Contudo ele apresentou uma faixa de 15% a mais de ocorrências que os valores empíricos.

O ano de 2015 apresentou nas estações inverno e primavera o maior número de ocorrências com 71% dos valores extremos identificados nesse ano, como é mostrado na Figura 28 (a). Todas as técnicas identificaram as mesmas estações que os valores empíricos, com o maior quantitativo de valores extremos. A técnica PAA apresentou um número bem abaixo de ocorrências, com menos de 25% do total de ocorrências do ano em estudo. Já o método PAA/RQ apresentou uma quantidade maior de ocorrências, contudo seu desempenho ainda foi considerado insuficiente, pois identificou menos de 36% das ocorrências de eventos extremos esperados. Com

relação ao APAA, o quantitativo total de ocorrências identificadas foi bem próximo dos valores empíricos, com cerca de 86% dos eventos extremos identificados por essa técnica.

Para o ano de 2016, as estações primavera, verão e inverno apresentaram o maior número de ocorrências com quase 90% do número total de eventos extremos desse ano, como é mostrado na Figura 28 (a). A estação da primavera foi responsável por quase 41% do total desses eventos. Todas as técnicas apresentaram as mesmas estações com o maior número de ocorrências que os valores empíricos. A técnica PAA identificou apenas 23% do quantitativo total de ocorrências para esse ano e para as estações com maiores índices de ocorrências de valores extremos. Já a técnica PAA/RQ identificou um número maior de ocorrências que o PAA, contudo ainda foi bem abaixo do esperado, com apenas 38% do total desse ano. A técnica APAA teve um desempenho mais satisfatório, identificando um número de ocorrências para cada estação bem próximo do esperado, contudo ele apresentou uma faixa de 19% a mais de ocorrências que os valores empíricos.

Figura 28 – Gráficos de múltiplas barras de ocorrências UP dos anos 2014, 2015 e 2016 para SERS com (a) Valores empíricos, (b) Técnica PAA/RQ, (c) Técnica PAA e (d) Técnica APAA



Fonte: Próprio Autor

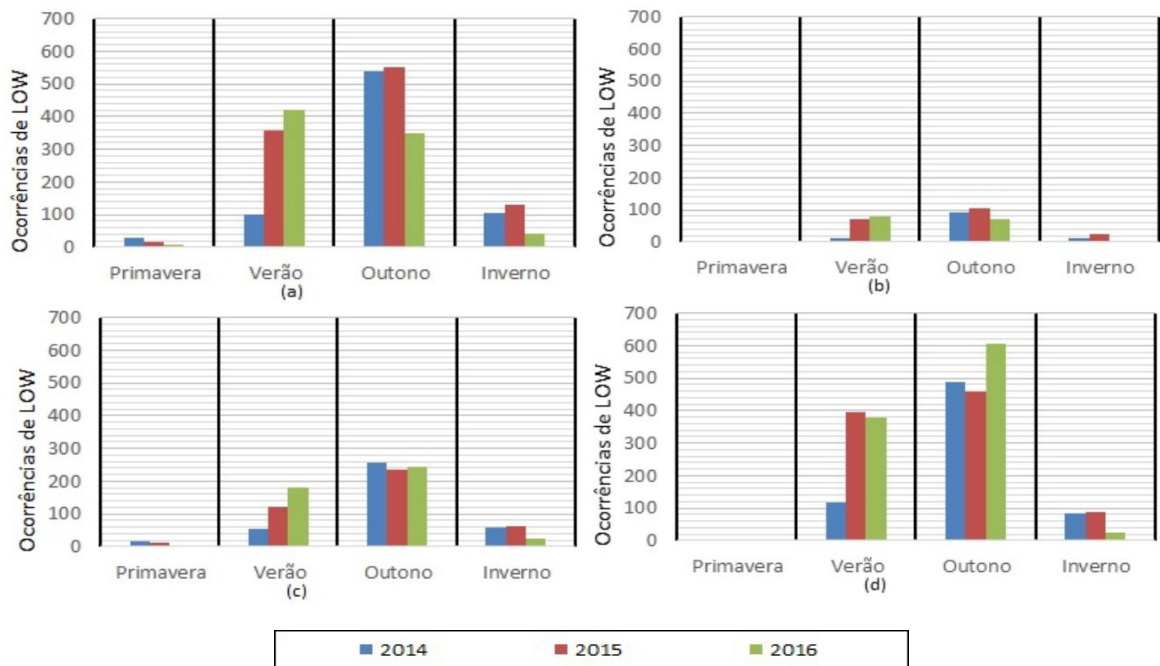
Para as ocorrências de LOW do SERNE, o ano de 2014 apresentou a maioria dos eventos extremos na estação outono com quase 70% das ocorrências, como pode ser observado na Figura 29 (a). Todas as técnicas identificaram a mesma estação do ano que os valores empíricos, com o maior número de ocorrências. Contudo, a técnica PAA apresentou um número bem abaixo do esperado, com apenas 16% do quantitativo total para o ano em estudo. Já a técnica PAA/RQ, teve um desempenho um pouco superior ao PAA; ele apresentou um pouco mais de 50% das ocorrências de eventos extremos no ano em questão. Com relação a técnica APAA,

ele apresentou um desempenho superior ao PAA/RQ, identificando quase 90% das ocorrências de eventos extremos do ano em questão e para as estações com maior índice de ocorrências de valores extremos.

Para o ano de 2015 as estações que apresentaram a maior parte das ocorrências são o verão e outono, como pode ser observado na Figura 29 (a), com 86% do quantitativo total do ano em questão. Todas as técnicas apresentaram as mesmas estações que os valores empíricos com o maior número de ocorrências. A técnica PAA identificou menos de 19% das ocorrências totais para o ano em estudo e para as estações com maiores índices de ocorrências de valores extremos. Já a técnica PAA/RQ apresentou um desempenho um pouco melhor que o PAA, pois identificou cerca de 41% das ocorrências de eventos extremos esperados para o ano analisado e 39% das ocorrências nas estações com maiores índices de eventos extremos. Para a técnica APAA, verificou-se um desempenho superior aos outros métodos estudados; ele apresentou cerca de 90% das ocorrências esperadas para o ano em análise e 94% dos eventos nas estações que apresentaram os maiores índices de ocorrências.

Para o ano de 2016 as estações verão e outono apresentaram o maior número de ocorrências com quase 94% do número total de eventos extremos desse ano, como pode ser observado na Figura 29 (a). Todas técnicas apresentaram as mesmas estações com o maior número de ocorrências que os valores empíricos. A técnica PAA identificou apenas 18% do quantitativo total de ocorrências para esse ano e quase 20% das ocorrências identificadas nas estações com maior quantitativo de valores extremos. Já a técnica PAA/RQ identificou um número maior de ocorrências que o PAA, contudo ainda foi bem abaixo do esperado, com apenas 55% do total desse ano. A técnica APAA teve um desempenho tão satisfatório quanto os outros anos, identificando um número de ocorrências para cada estação, com um quantitativo superior ao esperado, com 22% a mais de ocorrências. Tendo acontecido principalmente na estação do outono, com quase 73% a mais de ocorrências que os valores empíricos. Isso pode ter ocorrido devido esse mês apresentar na base de dados alguns valores repetidos de velocidade do vento, tornando uma janela maior que o necessário.

Figura 29 – Gráficos de múltiplas barras de ocorrências LOW dos anos 2014, 2015 e 2016 para SERNE com (a) Valores empíricos, (b) Técnica PAA/RQ, (c) Técnica PAA e (d) Técnica APAA



Fonte: Próprio Autor

5 CONCLUSÃO

As bases de dados utilizadas nesta dissertação são constituídas por dados de geração diária de vento, coletados a cada 10 minutos de 01 de janeiro de 2014 até 31 de dezembro de 2016. Elas correspondem às séries temporais de velocidade do vento de dois sítios eólicos localizados em diferentes regiões do Brasil, o SERS e SERNE situados na Região Sul e Nordeste respectivamente. Os algoritmos para análise desses dados foram implementadas na ferramenta MATLAB[®], estes permitem extrair informações sobre o comportamento cíclico e sazonal dessas séries, buscando realizar uma análise mais rápida que as convencionais presentes na literatura, assim como garantir uma análise eficiente e confiável dos bancos de dados. Os resultados iniciais obtidos são promissores, uma vez que permitem obter aspectos descritivos de maneira mais rápida sobre o comportamento dessas séries temporais devido à utilização de técnicas de redução de dimensionalidade das bases de dados. Essa análise é importante, pois assim é possível determinar os períodos em que as turbinas estão fora de operação, comprometendo a geração de energia dos parques eólicos, como também as áreas mais atrativas para a implantação de parques eólicos. A partir da análise do box-plot foi observado que os padrões de velocidade da região Sul apresentam valores de velocidade do vento mais altos que a região Nordeste, que não apresenta nenhum valor de UP. Contudo a região Nordeste apresenta faixas de velocidade do vento mais propícias ao funcionamento pleno da turbina eólica. Com relação as técnicas utilizados, a técnica PAA foi muito menos preciso, apresentando erros em relação à identificação das estações com o maior número de eventos extremos, e ainda a quantidade total de valores extremos foi muito menor do que o esperado. Os resultados para a técnica PAA/RQ foram ligeiramente melhores do que a técnica PAA, tanto em relação à identificação das estações que ocorrem a maioria dos valores extremos quanto em relação à quantidade de eventos detectados pela técnica. Esta técnica mostra melhor desempenho quando os eventos extremos não aconteceram sucessivamente, pois poderia ocorrer mais de um valor extremo na janela determinada pelo algoritmo PAA e, portanto, a regressão só identificaria o maior valor, nos casos em que as velocidades são maiores ou iguais a 20 m/s e o menor valor para o caso em que os valores de velocidade são inferiores ou iguais a 3 m/s. Já a técnica APAA foi o que mais se aproximou aos valores detectados empiricamente; ele também identificou de forma correta as estações do ano em que a maioria dos eventos extremos aconteceram. Comparando as técnicas PAA, PAA/RQ e APAA, observou-se que a técnica APAA teve o desempenho mais satisfatório, pois apresentou em todos os anos analisados um quantitativo de ocorrências de eventos extremos mais próximos dos valores esperados para os SERS e SERNE. Isso se deve ao fato de que esta técnica propõe uma análise da série temporal usando janelas deslizantes, para obter os segmentos que contêm a maioria dos valores extremos da velocidade do vento, modificando seu tamanho a partir das variações dos valores de dados analisados. Também é verificado que a maior parte das ocorrências de LOW nos três anos em análise para o SERS ocorrem no verão, inverno e outono, enquanto as ocorrências de UP ocorrem

na primavera, verão e inverno para os anos de 2014 e 2016, e na primavera e inverno para o ano de 2015, havendo um certo comprometimento do funcionamento das turbinas eólicas. Para o SERND, não foram verificadas nenhuma ocorrência de UP, no entanto, a maioria dos eventos extremos de LOW ocorrem no outono para o ano de 2014 e no outono e inverno para os anos de 2015 e 2016.

Como trabalhos futuros, é proposto o refinamento no uso do algoritmo APAA, com o objetivo de aumentar a sua eficiência, a partir da identificação de um quantitativo de valores extremos cada vez mais próximo dos valores esperados; também é proposto o uso de técnicas de mineração de dados para identificação de motivos (padrões de geração de energia, curto e longo prazos) (LIN et al., 2002), que forneçam informações de previsão de geração eólica a curto e médio prazo, já que esta fonte de energia é usada como complementação da energia gerada necessária para a carga instalada. Um outro ponto importante que pode ser explorado é a identificação de possíveis falhas no funcionamento da turbina eólica a partir da observação de valores discrepantes de velocidade do vento, geração real e geração prevista.

REFERÊNCIAS

ABEEOLICA. *Ventos: Tipos de vento, massas de ar, frentes frias e brisa marítima. De 11 de março de 2005*. 2005. Disponível em: <educacao.uol.com.br/disciplinas/ciencias/ventos-tipos-devento-massas-de-ar-frentes-frias-e-brisa-maritima.htm>. Acesso em: 19 de out. de 2017 as 14:38. Citado na página 23.

ABEEOLICA. *Associação Brasileira de Energia Eólica. Boletim n 5, de 11 de março de 2015*. 2017. Disponível em: <www.abeeolica.org.br/wp-content/uploads/2017/05/424_Boletim_Anual_de_Geracao_Eolica_2016_Alta.pdf>. Acesso em: 16 de jun. de 2017 as 21:51. Citado 2 vezes nas páginas 19 e 20.

AGRAWAL, R.; FALOUTSOS, C.; SWAMI, A. *Efficient similarity search in sequence databases: Foundations of data organization and algorithms*, pp. 69 - 84. [S.l.: s.n.], 1993. Citado na página 45.

ALBUQUERQUE, J. C. *Avaliação de curvas de potência em modelos de previsão de geração eólica em curto prazo*. Dissertação (Dissertação de Mestrado) — Programa de Pós Graduação em Engenharia Elétrica, Universidade Federal de Pernambuco, Recife, Brasil, 2015. Nenhuma citação no texto.

ALVES, A. *Análise do Desempenho de Rotores Eólicos de Eixo Horizontal*. Dissertação (Tese de Doutorado) — Programa de Pós Graduação em Engenharia Elétrica, Universidade Federal do Pará, Pará, Brasil, 1997. Nenhuma citação no texto.

AMARANTE, O. A. C. et al. Conceptual survey of generators and power electronics for wind turbines. *Centro de Pesquisas de Energia Elétrica/ CEPEL*, Brasília, Brasil, p. 5–91, 2001. Nenhuma citação no texto.

AMINIKHANGHAHI, S.; COOK, D. J. *A survey of methods for time series change point detection* Knowl. Inf. Syst., vol. 51, no. 2, p. 339–367, May 2017. Citado na página 48.

ARAO, M. O. P. *Estudo Comparativo de Sistemas Eólicos Utilizando Modelos Probabilísticos de Velocidade de Vento*. Dissertação (Dissertação de Mestrado) — Programa de Pós Graduação em Engenharia Mecânica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1989. Citado na página 30.

BARIONI, M. C. N. *Operações de consulta por similaridade em grandes bases de dados complexos*. Dissertação (Tese de Doutorado) — Programa de Pós Graduação em Matemática e Computação, Universidade de São Paulo, São Carlos, Brasil, 2006. Citado na página 36.

BONELLI, A. F. *Estudo do potencial eólico e previsão de ventos para geração de eletricidade em Santa Catarina*. Dissertação (Dissertação de Mestrado) — Programa de Pós Graduação em Engenharia Elétrica, Universidade Federal de Santa Catarina, Santa Catarina, Brasil, 2007. Nenhuma citação no texto.

BONELLI, A. F. *Modelagem e Simulação de Unidade Eólica para Estudos de Indicadores de Qualidade da Energia Elétrica*. Dissertação (Dissertação de Mestrado) — Programa de Pós Graduação em Engenharia Elétrica, Universidade Federal de Uberlândia, São Paulo, Brasil, 2010. Citado na página 35.

BOX, G. E. P. et al. *Time series analysis: forecasting and control*. [S.l.]: Editora John Wiley & Sons, 2008. Citado na página 39.

BUCHINSKY, M. Recent advances in quantile regression models: A practical guideline for empirical research. *The Journal of Humam Resources*, p. 88–126, 1998. Citado na página 50.

BURTON, T.; SHARPE, D.; BOSSANYI, E. *Wind Energy Handbook*. [S.l.]: John Wiley & Sons Ltd, 2001. Citado na página 34.

CARVALHO, B. C. *Desenvolvimento de Modelo Computacional de Sistemas Eólicos Utilizando Geradores Síncronos para Estudos de Desempenho no Contexto da Qualidade de Energia Elétrica*. Dissertação (Tese de Doutorado) — Programa de Pós Graduação em Engenharia Elétrica, Universidade Federal de Uberlândia, São Paulo, Brasil, 2006. Citado 3 vezes nas páginas 23, 25 e 34.

CARVALHO, F. W. *Ventos. De 11 de março de 2013*. 2013. Disponível em: <<http://www.brasilecola.com/geografia/vento.htm>>. Acesso em: 19 de out. de 2017 as 15:14. Citado na página 23.

CASTRO, R. M. G. *Introdução ergia Eólica*. [S.l.]: Instituto Superior Técnico, 2009. Nenhuma citação no texto.

CCBE Centro Brasileiro de Energia Eólica. 2013. Disponível em: <www.cresesb.cepel.br/principal.php>. Acesso em: 30 de mai. de 2017 as 18:14. Nenhuma citação no texto.

CECON, P. R. et al. *Métodos Estatísticos*. 1a edição. ed. Viçosa, Brasil: [s.n.], 2012. 229 p. Citado na página 50.

CHAKRABARTI, K. et al. Locally adaptive dimensionality reduction for indexing large time series databases. *ACM Transactions on Database Systems*, v. 7, p. 188–228, 2002. Citado 3 vezes nas páginas 20, 42 e 45.

CHAN, F. K. P.; FU, A. W. C.; YU, C. Haar wavelets for efficient similarity search of time-series: With and without time warping. *IEEE Transactions on Knowledge and Data Engineering*, p. 686–705, 2003. Citado na página 45.

CHIZI, B.; MAIMON, O. *Dimension Reduction and Feature Selection: Data mining and knowledge discovery handbook*, pp. 83 - 100. 2a edição. ed. [S.l.: s.n.], 2010. Citado na página 45.

CHU, S. et al. Iterative deepening dynamic time warping for time series. *Second SIAM International Conference on Data Mining*, p. 1–18, 2002. Citado 3 vezes nas páginas 43, 44 e 45.

COROTIS, R. B.; SIGI, A. B.; KLEIN, J. Probability models of wind velocity and magnitude and persistence. *Solar Energy*, v. 20, p. 134–143, 1978. Citado na página 29.

CUSTÓDIO, R. D. S. *Energia Eólica para a Produção de Energia Elétrica*. [S.l.]: Synergia, 2012. Citado 3 vezes nas páginas 26, 27 e 32.

DALMAZ, A. *Estudo do potencial eólico e previsão de ventos para geração de eletricidade em Santa Catarina*. Dissertação (Dissertação de Mestrado) — Programa de Pós Graduação em Engenharia Elétrica, Universidade Federal de Santa Catarina, Santa Catarina, Brasil, 2007. Nenhuma citação no texto.

DAMOUSIS, I. G.; ALEXIADIS, J. B.; DOKOPOULOS, P. S. A. Fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. *IEEE Transactions on energy conversion*, v. 19, 2004. Nenhuma citação no texto.

DANISH wind Industry Association. 2015. Disponível em: <www.windpower.org/en/tour.htm>. Acesso em: 30 de jan. de 2017 as 13:14. Nenhuma citação no texto.

DANISH Wind Turbine Manufacturers Association. 2010. Disponível em: <www.windpower.dk/>. Acesso em: 25 de set. de 2017 as 09:14. Citado 2 vezes nas páginas 28 e 29.

DAVENPORT, A. G. *The Relationship of Wind Structure to Wind Loading*. [S.l.]: National Physics Laboratory, 1963. Citado na página 30.

DIAS, B. H. *Programação Dinâmica Estocástica e Algoritmo de Fechos Convexos no Planejamento da Operação de Sistemas Hidrotérmicos*. Dissertação (Tese de Doutorado) — Programa de Pós Graduação em Engenharia Elétrica, PUC-Rio, Rio de Janeiro, Brasil, 1998. Nenhuma citação no texto.

DIEZ, P. F. *Energia Eólica*. Dissertação (Tese de Doutorado) — Departamento de ingeniería eléctrica y energética - Universidad de Cantabria, Cantabria, Espanha, 2003. Citado na página 33.

DINIZ, A. L.; SANTOS, T. M.; MACEIRA, M. E. P. Short term security constrained hydrothermal scheduling considering transmission losses. *IEEE/PES Transmission & Distribution Conference and Exposition: Latin America*, Caracas, Venezuela, 2006. Nenhuma citação no texto.

DRAPER, N.; SMITH, H. *Applied Regression Analysis*. [S.l.]: Wily-Interscience, 1998. Citado na página 50.

DUTRA, R.; MONTEZANO, B.; FERREIRA, J. C. E. *Energia Eólica - Princípios e Tecnologias: O recurso eólico: Capítulo 2*. 1. ed. Rio de Janeiro: [s.n.], 2009. Disponível em: <www.cresesb.cepel.br/download/tutorial/tutorial_eolica_2008_e-book.pdf>. Acesso em: 19 de out. de 2016 as 17:02. Citado na página 24.

EHLERS, R. *Análise de Séries Temporais*. Curitiba, Brasil: Laboratório de Estatística e Geoinformática, 2005. Citado na página 37.

ESTEBAN, M. D.; LÓPEZ, J. S.; NEGRO, V. Why offshore wind energy? *Journal of the World Renewable Energy Network*, v. 36, p. 444–450, 2004. Nenhuma citação no texto.

FADIGAS, E.; ARLINDO, P. *Energia Eólica: Caracterização do vento: Capítulo 5*. 1. ed. São Paulo: [s.n.], 2012. Disponível em: <www.enersud.com.br>. Acesso em: 19 de out. de 2016 as 17:52. Citado 4 vezes nas páginas 25, 31, 32 e 34.

FELGUEIRAS, A. Geração eólica. v. 2, p. 12–37, 2003. Citado 3 vezes nas páginas 25, 26 e 27.

FERREIRA, A. A. et al. Adaptive Piecewise and Symbolic Aggregate Approximation as an Improved Representation Method for Heat Waves Detection. *Computing Conference - IEEE*, London, UK, 2018. Citado na página 48.

FERRERO, C. A. *Algoritmo kNN para previsão de dados temporais: funções de previsão e critérios de seleção de vizinhos próximos aplicados a variáveis ambientais em limnologia*. Dissertação (Dissertação de Mestrado) — Universidade de São Paulo - USP, São Paulo, Brasil, 2009. Citado 3 vezes nas páginas 37, 38 e 40.

FINK, E. *Data Mining in Time Series Databases*: Capítulo indexing of compressed time series, pp. 43 - 65, 2004. Citado na página 42.

FRANCESCHETTI, N. N.; SIMS, M. G.; BOSE, B. K. Otimização de um sistema de geração de energia eólica através de controle fuzzy. *SBA Controle & Automação*, v. 10, p. 48–58, 1999. Nenhuma citação no texto.

GASCH, R.; TWELE, J. *Wind Power Plants: Fundamentals, Design, Construction and Operation*. [S.l.]: Germany textbook, 2012. Nenhuma citação no texto.

GNEITING, T. Quantiles as optimal point forecasts. *International Journal of Forecasting*, p. 197–207, 2011a. Nenhuma citação no texto.

GORENSTIN, B. et al. Stochastic optimization of a hydro-thermal system including network constraints. *IEEE Transactions on Power Systems*, p. 791–797, 2009. Nenhuma citação no texto.

GUJARATI, D. N. *Econometrica básica: Econometria*. 5a edição. ed. São Paulo, Brasil: [s.n.], 2011. 920 p. Citado na página 50.

GWEC. *Global Wind Energy Council. Opening up new markets for business: Mongócia n 8, de 5 de maio de 2016*. 2016. 8TH Nacional Renewable Energy Forum. Disponível em: <<http://gwec.net/publications/global-wind-report-2/>>. Acesso em: 18 de jul. de 2017 as 8:50. Citado 2 vezes nas páginas 18 e 19.

HAN, J.; KAMBER, M. *Data Mining: Concepts and Techniques*. 2a edição. ed. [S.l.]: Editora Morgan Kaufman, 2006. Citado 2 vezes nas páginas 36 e 39.

HAN, J.; KAMBER, M.; PEI, J. *Data mining: Concepts and techniques: The morgan kaufmann series in data management systems*. [S.l.]: Editora Elsevier Science, 2011. Citado 2 vezes nas páginas 41 e 45.

HANSEN, A.; HELLE, L.; BLAABJERG, e. a. Conceptual survey of generators and power electronics for wind turbines. *Riso National Laboratory*, p. 5–91, 2001. Nenhuma citação no texto.

HAO, L.; NAIMAN, D. Q. *Quantile Regression*. [S.l.]: Sage Publications, 2007. 126 p. Citado 3 vezes nas páginas 50, 52 e 53.

HAU, E. *Wind Turbines Fundamentals, Technologies, Application, Economics*. [S.l.]: Editora Springer, 2006. Nenhuma citação no texto.

HENNESSEY, J. P. Some aspects of wind power statistics. *J. Appl. Meteor.*, v. 16, p. 119–128, 1976. Citado 2 vezes nas páginas 30 e 31.

HERING, A.; GENTON, M. G. Powering up with space-time wind forecasting. *Journal of the American Statistical Association*, p. 94–104, 2010. Nenhuma citação no texto.

HERRERA, M. et al. Quantile regression ensemble for summer temperatures time series and its impact on built environment studies. *IEMSS Proceedings of International Environmental Modelling and Software Society*, v. 7, p. 602609, 2016. Citado na página 20.

- JUSTUS, C. G.; HARGRAVES, W. R.; GRADER, D. Methods for estimating wind speed frequency distributions. *Journal of Applied Meteorology*, v. 17, p. 350–353, 1978. Citado na página 30.
- JUSTUS, C. G.; HARGRAVES, W. R.; YALCIN, A. Nationwide assessment of potential output from wind powered generators. *J. Appl. Meteor.*, v. 15, p. 673–678, 1976. Citado na página 30.
- JUSTUS, C. G.; MIKHAIL, A. Height variation of wind speed and wind distribution. *Geophysical Research Letters*, v. 03, p. 261–264, 1976. Citado na página 30.
- KADOWAKI, M. *Modelo de Programação da Operação de Sistemas Hidrotérmicos de Curto-Prazo*. Dissertação (Tese de Doutorado) — Programa de Pós Graduação em Engenharia Elétrica e Computação, Universidade Estadual de Campinas, Campinas, Brasil, 2012. Nenhuma citação no texto.
- KEOGH, E. et al. Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and Information Systems, Springer-Verlag London Limited*, v. 3, p. 263286, 2001. Citado 4 vezes nas páginas 20, 46, 47 e 49.
- KEOGH, E.; KASETTY, S. On the need for time series data mining benchmarks: a survey and empirical demonstration. *8th International Conference on Knowledge Discovery and Data Mining*, p. 102110 and 349 – 371, 2002. Citado na página 42.
- KEOGH, E.; KASETTY, S. On the need for time series data mining benchmarks: a survey and empirical demonstration. *IEEE Transactions on Power Systems*, 2003. Citado 2 vezes nas páginas 36 e 41.
- KEOGH, E.; LONARDI, S.; RATANAMAHATANA, C. A. Towards parameter-free data mining. *10th International Conference on Knowledge Discovery and Data Mining*, New York, USA, p. 206215, 2004. Citado na página 42.
- KEOGH, E. et al. Compression-based data mining of sequential data. *Data Mining and Knowledge Discovery*, p. 99129, 2007. Citado na página 42.
- KOENKER, R. W.; BASSET, G. J. *Regression quantiles*: *Econometria*, v. 50, p. 43-61, 1978. Citado 3 vezes nas páginas 49, 50 e 52.
- KOENKER, R. W.; BASSET, G. J. *Robust tests for heteroscedasticity based on regression quantiles*: *Econometria*. [S.l.: s.n.], 1982. Citado na página 51.
- KOENKER, R. W.; MACHADO, J. Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, p. 1296–1310, 1999. Citado na página 51.
- KORN, F.; JAGADISH, H. V.; FALOUTSOS, C. Efficiently supporting ad hoc queries in large datasets of time sequences. *ACM SIGMOD international conference on Management of data, SIGMOD 97*, New York, USA, p. 289–300, 2013. Citado na página 45.
- LEMOS, D. F. A. *Análise de Projeto de Turbina Eólica de grande porte para as condições climáticas da Região Nordeste do Brasil*. Dissertação (Dissertação de Mestrado) — Programa de Pós Graduação em Engenharia Elétrica, Universidade Federal de Pernambuco, Recife, Brasil, 2005. Citado na página 19.

LIN, J.; KEOGH, E. Clustering of time-series subsequences is meaningless: implications for previous and future research. *Knowledge and Information Systems*, v. 8, p. 154177, 2005. Citado na página 41.

LIN, J. et al. Experiencing sax: a novel symbolic representation of time series. *Data Mining and Knowledge Discovery*, p. 107144, 2007. Citado na página 47.

LIN, J. et al. Finding motifs in time series. in proceedings of the 2nd workshop on temporal data mining. *8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p. 11, 2002. Citado 2 vezes nas páginas 46 e 72.

LOPES, A. M. *Produção Eólica e Enquadramento Técnico-Económico em Portugal*. Dissertação (Dissertação de Mestrado) — Faculdade de Engenharia da Universidade do Porto, Porto, Portugal, 2007. Nenhuma citação no texto.

MACCARINI, M. C. *Inversor Monofásico Sincronizado para a Conexão de um Gerador Eólico de Elétrica: Estudo, Projeto e Implementação*. Dissertação (Dissertação de Mestrado) — Programa de Pós Graduação em Engenharia Elétrica, Universidade Federal de Santa Catarina, Santa Catarina, Brasil, 2009. Citado 2 vezes nas páginas 28 e 34.

MALETZKE, A. G. e. a. *Time series classification with motifs and characteristics: Soft computing for business intelligence*: pages 125 - 138. [S.l.]: Editora Springer Berlin Heidelberg, 2014. ISBN 978-3-642-53736-3. Citado na página 37.

MALI, S. S.; KUSHARE, B. E. Mppt algorithms: Extracting maximum power from wind turbines. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, 2013. Nenhuma citação no texto.

MANWELL, J. F.; MCGOWAN, J. G.; ROGERS, A. L. *Wind Energy Explained: Theory, Design and Application*. [S.l.]: Editora Wiley, 2010. Nenhuma citação no texto.

MICHALSKI, R. S.; KUBAT, M. *Machine learning and data mining* [S.l.]: Editora John Wiley & Sons, 1998. Citado na página 39.

MORCHEN, F. *Time series knowledge mining*. Dissertação (Phd Thesis) — Department of Mathematics and Computer Science Philipps-University, Hesse, Germany, 2013. Citado 2 vezes nas páginas 40 e 42.

MORETTIN, P. A.; TOLOI, C. M. C. *Análise de Series Temporais*. [S.l.]: Editora Edgard Blucher, 2004. Citado 4 vezes nas páginas 37, 38, 39 e 40.

MULJADI, E.; BUTTERFIELD, C. Pitch- controlled variable-speed wind turbine generation. *IEEE Transactions on Industry Applications*, v. 37, p. 240246, 2001. Nenhuma citação no texto.

NETO, A. S. *Análise e Controle de Centrais Eólicas a Velocidade Variável Utilizando ATP-Draw*. Dissertação (Dissertação de Mestrado) — Programa de Pós Graduação em Engenharia Elétrica, Universidade Federal de Pernambuco, Recife, Brasil, 2009. Citado na página 34.

NEYMAN, J.; PEARSON, E. *On the use and interpretation of certain teste criteria for purposes of statistical inference*: *Biometrika*, v. 20, p. 175-240., 1928. Citado 2 vezes nas páginas 51 e 52.

NUNES, M. V. A. *Avaliação do comportamento de aerogeradores de velocidade fixa e variável integrados em redes elétricas fracas*. Dissertação (Tese de Doutorado) — Programa de Pós Graduação em Engenharia Elétrica, Universidade Federal de Santa Catarina, Santa Catarina, Brasil, 2003. Nenhuma citação no texto.

OLIVEIRA, R. G. *Controle e Integração a Rede Elétrica de um Sistema Eólico com Gerador de Indução Duplamente Excitado*. Dissertação (Tese de Doutorado) — Programa de Pós Graduação em Engenharia Elétrica, Universidade Federal de Minas Gerais, Minas Gerais, Brasil, 2004. Nenhuma citação no texto.

RABINER, L. A tutorial on hidden markov models and selected applications in speech recognition. *IEEE*, p. 257–286, 1989. Citado na página 46.

RAM, G. N.; KIRUTHIGA, A.; SHREE, J. D. A novel maximum power point tracking system for wind-energy-conversion system using particle swarm optimization. *IJERT International Journal of Engineering Research & Technology*, 2014. Nenhuma citação no texto.

RATANAMAHATANA, C. A.; KEOGH, E. Making time-series classification more accurate using learned constraints. *SIAM International Conference on Data Mining*, p. 11–24, 2004. Citado na página 44.

RATANAMAHATANA, C. A. et al. A novel bit level time series representation with implication of similarity search and clustering. p. 771–777, 2005. Citado na página 46.

REGUSE, W. *Panorama Atual da Utilização da Energia Elétrica: In: Fontes não-convencionais de energia: as tecnologias solar e eólica e de biomassa*. [S.l.]: Editora UFSC, 1999. Nenhuma citação no texto.

RIOS, R. *Improving time series modeling by decomposing and analyzing stochastic and deterministic influences*. Dissertação (Tese de Doutorado) — Programa de Pós Graduação em Matemática e Computação, Universidade de São Paulo, São Carlos, São Paulo, Brasil, 2013. Citado na página 39.

ROHATGI, J. S.; NELSON, V. *Wind Characteristics - An Analysis for the Generation of Wind Power*. [S.l.]: Alternative Energy Institute, 1994. Citado na página 30.

ROSAS, P. A. C.; ESTANQUEIRO, A. I. *Guia de Projeto Elétrico de Centrais Eólicas*. CBEE - Centro Brasileiro de Energia Eólica, março de 2003. 2003. Disponível em: <energiasrenovaveis.com/images/upload/Guia_Eletrico_CBEE_eolica.pdf>. Acesso em: 16 de jun. de 2017 as 21:51. Nenhuma citação no texto.

SANSIGOLO, C. A. Distribuição de probabilidade de velocidade e potia do vento. *Revista Brasileira de Meteorologia*, v. 02, p. 207–214, 2005. Citado na página 30.

SORENSE, P.; HANSEN, A.; JANOSI, L. e. a. Simulation of interaction between wind farm and power system. *Ris National Laboratory*, p. 355, 2001. Nenhuma citação no texto.

TAYLOR, J. W.; MCSHARRY, P. E.; BUIZZA, R. Wind power density forecasting using ensemble predictions and time series models. *IEEE Transactions on Energy Conversion*, p. 775–782, 2009. Nenhuma citação no texto.

THOMANN, G. C.; BARFIELD, M. J. The time variation of wind speeds and windfarm power output in kansas. *IEEE Transactions on Energy Conversion*, v. 3, p. 44–49, 1998. Citado na página 29.

THONGAM, J. S. et al. Wind speed sensorless maximum power point tracking control of variable speed wind energy conversion systems. *Electric Machines and Drives Conference*, 2009. Nenhuma citação no texto.

THONGAM, J. S.; OUHROUCHE, M. *Fundamental and Advanced Topics in Wind Power: Mppt control methods in wind energy conversion systems: Capítulo 15*. [S.l.]: Editora Intech, 2011. Nenhuma citação no texto.

TROEN, I.; PETERSEN, E. L. *European Wind Atlas*. [S.l.]: Riso National Laboratory, 1989. Citado na página 30.

VLACHOS, M.; GUNOPULOS, D. *Data Mining in Time Series Databases: Capítulo indexing time-series under conditions of noise*, pp. 67 - 100, 2004. Citado na página 42.

VRIES, E. D. Where to next? developments and trends in wind turbines. *Renewable ENERGY World*, p. 6371, 2002. Nenhuma citação no texto.

WANG, X. et al. Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery*, p. 275–309, 2013. Citado na página 45.

WIND Turbine Power Output Variation with Steady Wind Speed. 2014. Disponível em: <www.wind-power-program.com/turbine_characteristics.htm.> Acesso em: 30 de set. de 2017 as 12:14. Citado na página 33.

XAVIER, L.; COSTA, F.; MACEIRA, M. Aprimoramento da modelagem da função de produção energética das usinas hidroelétricas no modelo decomp: Metodologia e resultados. *XVIII SNPTEEE - Seminário Nacional de Produção e Transmissão de Energia Elétrica*, Curitiba, Brasil, 2005. Nenhuma citação no texto.