Universidade Federal de Pernambuco
Centro de Tecnologia e Geociências
Departamento de Engenharia Mecânica

Pós-graduação em Engenharia Mecânica

Detecção e Diagnóstico de Falhas na Performance de Aerogeradores

Guilherme Tenório Maciel da Cunha Pedrosa

Dissertação de Mestrado

Recife
05 de Setembro de 2016

Universidade Federal de Pernambuco
Centro de Tecnologia e Geociências
Departamento de Engenharia Mecânica

Guilherme Tenório Maciel da Cunha Pedrosa

Detecção e Diagnóstico de Falhas na Performance de Aerogeradores

Trabalho apresentado ao Programa de Pós-graduação em Engenharia Mecânica do Departamento de Engenharia Mecânica da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Mestre em Engenharia Mecânica.

Orientador: *Pedro André Carvalho Rosas*
Co-orientador: *Alexandre Carlos Araújo da Costa*

Recife
05 de Setembro de 2016

Catalogação na fonte
Bibliotecária Rosineide M. G. Luz, CRB-4 / 1361

P372d Pedrosa, Guilherme Tenório Maciel da Cunha.
Detecção e diagnóstico de falhas na performance de aerogeradores / Guilherme Tenório Maciel da Cunha Pedrosa - 2016.
68 folhas, Il.; Tabs.; Abr. e Sigl.

Orientador: Prof. Dr. Pedro André Carvalho Rosas.
Coorientador: Prof. Dr. Alexandre Carlos Araújo da Costa.

Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia Mecânica, 2016.
Inclui Referências.

1. Engenharia Mecânica. 2. Energia eólica. 3. Monitoramento de performance. 4. Detecção de falhas. 5. Diagnóstico de falhas. 6. Aerogeradores. I. Rosas. Pedro André Carvalho(Orientador). II. Costa, Alexandre Carlos Araújo da(Coorientador). III.Título.

UFPE

621 CDD (22. ed.) BCTG/2017 – 18

05 de setembro de 2016.

“DETECÇÃO E DIAGNÓSTICO DE FALHAS NA PERFORMANCE DE AEROGERADORES”

GUILHERME TENÓRIO MACIEL DA CUNHA PEDROSA

ESTA DISSERTAÇÃO FOI JULGADA ADEQUADA PARA OBTENÇÃO DO TÍTULO DE MESTRE EM ENGENHARIA MECÂNICA

ÁREA DE CONCENTRAÇÃO: PROCESSOS E SISTEMAS TÉRMICOS

APROVADA EM SUA FORMA FINAL PELO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA/CTG/EEP/UFPE

_____
Prof. Dr. PEDRO ANDRÉ CARVALHO ROSAS
ORIENTADOR/PRESIDENTE

_____
Prof. Dr. ALEXANDRE CARLOS ARAÚJO DA COSTA
COORIENTADOR

_____
Prof. Dr. CEZAR HENRIQUE GONZALEZ
COORDENADOR DO PROGRAMA

BANCA EXAMINADORA:

_____
Prof. Dr. PEDRO ANDRÉ CARVALHO ROSAS (UFPE)

_____
Prof. Dr. ALEXANDRE CARLOS ARAÚJO DA COSTA (UFPE)

_____
Prof. Dr. ALEX MAURÍCIO ARAÚJO (UFPE)

_____
Prof. Dr. CARLOS COSTA DANTAS (UFPE)

Aos meus entes queridos

Agradecimentos

Agradeço a todos que contribuiram para tornar este trabalho possível.

A little learning is a dangerous thing. Drink deep, or taste not the Pierian spring; There shallow draughts intoxicates the brain, and drinking largely sobers us again.

—ALEXANDER POPE (An Essay on Criticism, 1709)

Resumo

Aerogeradores modernos são capazes de coletar e armazenar uma enorme quantidade de informações graças ao sistema de controle supervisório e aquisição de dados (SCADA) neles comumente embarcados. Esses dados são, porém, muitas vezes negligenciados em lugar de empregados em prol da otimização das tarefas de operação, manutenção e maximização de performance das centrais eólicas. Com vistas a explorar ao máximo as informações provenientes do sistema SCADA, este trabalho estabelece um sistema de detecção e diagnóstico de falhas relativas à potência ativa do aerogerador. Desvios de potência são detectados por meio da comparação entre dados observacionais da potência ativa e saídas de modelos uni e multivariáveis. Caso detectados, os desvios na potência são submetidos a um método de diagnóstico inovador baseado na informação mútua entre variáveis do SCADA e os próprios desvios. Tal método tem como objetivo indicar as variáveis do SCADA que melhor explicam os desvios de potência ativa. O sistema de detecção e diagnóstico de falhas aqui proposto foi aplicado a dois estudos de caso utilizando-se de dados observacionais de aerogeradores reais de uma dada central eólica, cedidos em regime de confidencialidade pela empresa proprietária da central. A validação dos modelos se dá através da comparação entre seus resultados e os metadados do livro de ocorrências da central eólica. É observado que o método de detecção de falhas foi capaz de identificar até 100% dos dados com falha para um dos casos apresentados, apesar de apresentar dificuldades na detecção de pequenas variações de potência no segundo caso, detectando apenas $47,18\%$ dos dados com falha. Contudo, o novo método de diagnóstico se mostrou eficaz em apontar as variáveis com maior possibilidade de explicar as variações da potência, mesmo para falhas similares. As variáveis que se destacaram coincidiram com a causa raiz dos problemas apresentados segundo o livro de ocorrências da central eólica.

Palavras-chave: Energia eólica. Monitoramento de performance. Detecção de falhas. Diagnóstico de falhas. Aerogeradores. Sistema SCADA.

Abstract

Wind turbines are capable of collecting and storing a great amount of information due to the supervisory control and data acquisition systems (SCADA) usually boarded in modern machines. These data, however, are neglected in most cases rather than employed to optimize the mainenance and operation tasks and maximize the performance of wind farms. Aiming to exploit SCADA data information, this work establishes a wind turbine performance fault detection and diagnosis system, with respect to the wind turbine generator active power. Power deviation is detected through comparison between observational active power data and the output of uni and multivariable models. If detected, power deviation is subjected to an innovative fault diagnosis method based on the mutual information between itself and the SCADA data. This method aims to identify which SCADA variables best explain active power deviations. The fault detection and diagnosis system hereby proposed are applied to two case studies based on real wind turbine generator data of a specific wind farm, given under confidential clauses. It is observed that the failure detection method was capable of detecting 100% of the data with failure for one of the study cases, even though it presented difficulties to detect the failure for smaller residual variations, resulting in a $47,18\%$ detection rate. Nonetheless, the new diagnosis method showed itself efficient in pointing SCADA variables prone to explain active power deviation, even for similar failure modes. The SCADA variables indicated the failure root cause according to the wind farm metadata logbook.

Keywords: Wind energy. Performance monitoring. Failure detection. Failure diagnosis. Wind turbine generators. SCADA system.

Lista de Figuras

Figura 1: Capacidade instalada cumulativa global 2000-2015. 15

Figura 2: Coeficiente de potência C_p em função da velocidade de ponta do rotor λ para diferentes modelos de turbinas eólicas. 19

Figura 3: Componentes de um aerogerador. 20

Figura 4: Arquitetura de um sistema SCADA 21

Figura 5: Curva de potência 22

Figura 6: Diagrama de caixas 23

Figura 7: Custo total versus taxa de falha de componentes principais nos primeiros 5 anos de operação de parques eólicos da DNV-GL 26

Figura 8: Métodos para elaboração de curva de potência. 27

Figura 9: Fluxograma com o caminho percorrido pelos dados SCADA até o resultado do diagnóstico do aerogerador. 30

Figura 10: Exemplo de sinal bidimensional com presença de ruído. 34

Figura 11: Relação entre a entropia, entropia condicional, entropia conjunta e informação mútua de duas variáveis aleatórias. 36

Figura 12: Zonas de falhas distintas de uma curva de potência. 38

Figura 13: Esquema de um neurônio artificial. 39

Figura 14: Arquitetura de rede neural artificial MLP. 40

Figura 15: Janela móvel aplicada com um *offset* de $dt = 144$ pontos e janela de comprimento $w = 4320$ pontos. 42

Figura 16: Erro RMS por bin de velocidade de 1m/s para a diferença entre cada um dos modelos considerados e a potência observada na amostra S2 para o caso 1. 47

Figura 17: Erro RMS por bin de velocidade de 1m/s para a diferença entre cada um dos modelos considerados e a potência observada na amostra S1 (falha). 48

Figura 18: Gráfico do resíduo normalizado das amostras S1 pela velocidade de vento (m/s), com as bandas de falha superior e inferior por bin de velocidade de 1m/s. 50

Figura 19: Caso 1 - Gráfico superior: informação mútua entre o resíduo condicionado e variáveis do SCADA com destaque para as variáveis mais prováveis de serem a causa raiz da falha. Gráfico inferior: resíduo entre o modelo P_9 de CP e os dados observacionais de potência ativa com regulações detectadas e não detectadas. 52

Figura 20: Caso 2 - Gráfico superior: informação mútua entre o resíduo condicionado e variáveis do SCADA com destaque para as variáveis mais prováveis de serem a causa raiz da falha. Gráfico inferior: resíduo entre o modelo *NN* de CP e os dados observacionais de potência ativa com regulações detectadas e não detectadas indicadas. 53

Figura 21: Distorções da CP modelada utilizando dados provenientes da PCA para o caso 1. 64

Figura 22: Distorções da CP modelada utilizando dados provenientes da IM para o caso 1. 64

Figura 23: CP polinomial grau 5 e grau 9 para o caso 1. 65

Figura 24: CP do fabricante e NN para o caso 1. 65

Figura 25: Distorções da CP modelada utilizando dados provenientes da PCA para o caso 2. 66

Figura 26: Distorções da CP modelada utilizando dados provenientes da IM para o caso 2. 66

Figura 27: CP polinomial grau 5 e grau 9 para o caso 2. 67

Figura 28: CP do fabricante e NN para o caso 2. 67

Lista de Tabelas

Tabela 1: Modelos de CP e dados de entrada considerados no trabalho. 37

Tabela 2: Quantidade de segmentações por zonas de velocidade para modelo linear segmentado de CP. 38

Tabela 3: Porcentagem de dados potencialmente anômalos considerando a velocidade, a potência ativa e ambos, de acordo com metodologia de garantia de qualidade dos dados. 44

Tabela 4: Variáveis selecionadas a partir do ranking da Informação Mútua entre as variáveis do SCADA e Potência Ativa. 46

Tabela 5: Cruzamento de dados do sistema de detecção de falhas com o de garantia de qualidade, para o modelo de CP polinomial e do fabricante. 51

Tabela 6: Lista de variáveis do sistema SCADA filtradas. 62

Lista de Abreviaturas e Siglas

CMS	*Condition Monitoring System*
CP	Curva de Potência
DNV-GL	*Det Norske Veritas and Germanischer Lloyd*
EAP	Estimativa anual de produção
GWEC	*Global Wind Energy Council*
IM	Informação Mútua
IQR	Intervalo interquartílico
IRENA	*International Renewable Energy Agency*
LCOE	*Levelized Cost of Energy*
LM	*Levenberg-Marquardt*
LS	Linear Segmentado
MLP	*Multi layer perceptron*
NARX	*Nonlinear autoregressive exogenous*
GMR	*Generalized mapping regressor*
GRNN	*General regression neural network*
NN	*Neural Networks*
O&M	Operação e manutenção
PCA	*Principal Componenet Analysis*
P_{man}	Potência garantida pelo fabricante do aerogerador
P_n	Polinômio de grau n
P_{nom}	Potência nominal
Q_i	Quartil i
RMS	*Root Mean Squared*
S1	Conjunto amostral contendo a falha estudada
S2	Conjunto amostral livre da falha estudada
SCADA	*Supervisory Control and Data Acquisition*
ST	Conjunto amostral de treino
v_{cut-in}	Velocidade de arranque
$v_{cut-out}$	Velocidade de corte do aerogerador
v_{rated}	Velocidade nominal do aerogerador

Sumário

1 Introdução **15**

1.1 Motivação 15

1.2 Objetivos 16

2 Conceitos Preliminares **18**

2.1 O Aerogerador 18

2.2 Sistemas SCADA 20

2.3 Curvas de Potência 21

2.4 Modelos de Manutenção 22

2.5 Diagrama de Caixa 23

3 Revisão Bibliográfica **25**

3.1 Monitoramento de Performance de Aerogeradores 25

3.2 Enunciado do Problema 28

4 Metodologia e Modelos **30**

4.1 Metodologia 30

4.2 Modelos 31

4.2.1 Garantia de Qualidade 31

4.2.2 Redução de Dimensionalidade 32

4.2.2.1 Análise de Componentes Principais 33

4.2.2.2 Informação Mútua 35

4.2.3 Detecção de Falhas 36

4.2.3.1 Regressão Multilinear baseada nos Mínimos Quadrados . . . 37

4.2.3.2 Redes Neurais Artificiais 39

4.2.4 Diagnóstico de Falhas 41

5 Estudos de Caso **43**

5.1 Descrição dos Dados 43

5.2 Resultados e Discussões 44

5.2.1 Garantia de Qualidade 44

5.2.2 Redução de Dimensionalidade 45

5.2.3 Detecção de Falhas 46

5.2.3.1 Modelagem de Curva de Potência 46

5.2.3.2 Estabelecimento de Limiares de Falha do Resíduo 49

5.2.4 Diagnóstico de Falhas 51

6 Conclusões e Perspectivas **54**

6.1 Conclusões 54

6.2 Recomendações 55

Referências Bibliográficas **56**

Apêndice A Variáveis do Sistema SCADA **62**

Apêndice B Curvas de Potência **64**

B.1 Caso 1 64

B.2 Caso 2 66

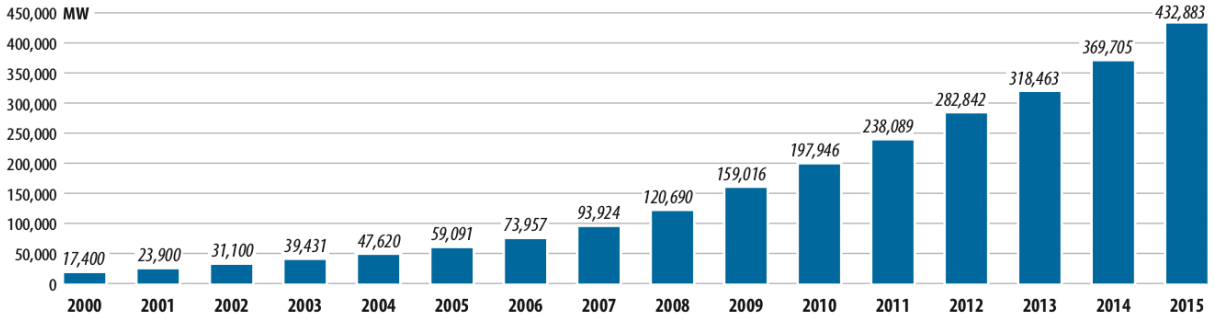
1 Introdução

1.1 Motivação

O ano de 2014 foi o primeiro ano em que as emissões de gás carbônico permaneceram estáveis, mesmo com o aumento do consumo energético global. Tal fenômeno se explica pelo aumento da inserção e eficiência das energias renováveis, que só vêm aumentando. Dentre as fontes de energia renováveis que impulsionaram o rápido crescimento da capacidade instalada, em primeiro lugar encontra-se a energia eólica (REN21, 2015).

Segundo o *Global Wind Energy Council* (GWEC, 2015), mais de 63 GW foram instalados em 2015, o que representa um aumento de mais de 23% com relação ao ano anterior. Ao final de 2015, a capacidade instalada global atingiu o valor de 432,88 GW, confirmando mais uma vez sua tendência de rápido aumento, conforme apresentado na Figura 1. No Brasil, este valor foi de 8,72 GW, liderança absoluta na América Latina sendo o único país na região com capacidade instalada superior a 1 GW.

Figura 1 – Capacidade instalada cumulativa global 2000-2015.



Fonte – GWEC (2015)

Tal crescimento do número de aerogeradores no mercado tem se dado observando critérios contratuais rígidos. Por exemplo, as centrais eólicas devem se comprometer com uma disponibilidade contratual quanto à geração acima de 98%. Para atingir tal valor, as centrais demandam soluções de monitoramento e manutenção contínuas (NILSSON; BERTLING, 2007).

Em seu estudo, Hughes (2012) mostrou que, apesar do relativo domínio sobre a tecnologia de aerogeradores onshore, o fator de capacidade dos aerogeradores na Grã Bretanha e Dinamarca tem caído significativamente com o passar dos anos. Como a rentabilidade dos empreendimentos

eólicos está diretamente relacionada ao custo nivelado de energia (LCOE[1]), efetivamente tem-se uma queda na lucratividade do negócio com o tempo.

Atualmente, aerogeradores modernos já vêm embarcados com um sistema supervisório de aquisição de dados (SCADA[2]), que constitui um importante meio de controle e fonte de informações. O fato de já estar presente, torna o sistema SCADA um forte candidato a ferramenta para monitoramento de baixo custo, pois não há despesas com sua instalação. Trabalhos com foco no monitoramento de performance utilizando seus dados já foram propostos (KUSIAK; VERMA, 2013; BUTLER; RINGWOOD; O'CONNOR, 2013; ULUYOL et al., 2011; ULUYOL; PARTHASARATHY, 2012). Contudo, não existe um consenso sobre as técnicas e metodologias para tal finalidade (WANG; INFIELD, 2013). Além disso, os autores focaram seus esforços em um número reduzido de variáveis, das centenas disponibilizadas pelos sistemas SCADA. É nesse contexto, que se pretende explorar o potencial de utilização de um maior número de variáveis do SCADA para o monitoramento de performance.

1.2 Objetivos

É proposta, a elaboração de um sistema de monitoramento automático com o foco na detecção e diagnóstico de falhas na performance de aerogeradores. O sistema é composto por quatro módulos com as seguintes funções e objetivos:

- **Garantia de qualidade:** módulo responsável por verificar a integridade dos dados de entrada do modelo quanto à sua consistência, erros ocasionados na aquisição, armazenamento e/ou manipulação. Será realizada a adaptação do trabalho de Moraes (2014) de garantia de qualidade para dados observacionais de vento em superfície de forma a que se aplique a dados de vento e potência dos aerogeradores;

- **Redução de dimensionalidade:** etapa em que as informações disponíveis de diferentes variáveis do sistema SCADA são consideradas de forma a selecionar ou reduzir o seu conjunto inicial apenas àquelas relevantes para explicar o comportamento da variável de interesse. Será avaliada a eficácia da utilização de técnicas de redução de dimensão na elaboração de modelos de curva de potência mais fidedignos à real condição de operação da máquina;

- **Detecção de falhas:** modelos de curva de potência são elaborados com base nos resultados da redução de dimensionalidade. A potência observada é, então, comparada à potência

[1] Custo nivelado de energia (do inglês *levelized cost of energy*): parâmetro resultante da divisão dos gastos com a implantação, operação e manutenção de uma central eólica por sua produção em quilowatt-hora gerados durante toda sua operação.

[2] *Supervisory Control and Data Acquisition*

estimada do melhor modelo, gerando-se um resíduo. O método não paramétrico do diagrama de caixas (*box-whiskers*) é aplicado ao resíduo para a determinação dos limiares de falha para diferentes faixas de velocidade do vento. O trabalho iniciado por Uluyol et al. (2011) será aperfeiçoado através da utilização do *boxwhisker* não paramétrico com bandas assimétricas para detecção de desvios na potência ativa do aerogerador;

- **Diagnóstico de falhas:** há identificação das variáveis capazes de explicar os desvios da potência ativa com respeito aos valores esperados. Para isto serão aplicados métodos baseados na teoria da informação.

Os capítulos seguintes focam em fundamentar e validar a metodologia proposta. Eles estão dispostos da seguinte maneira: o Capítulo 2 apresenta conceitos utilizados recorrentemente e necessários para melhor compreensão e contextualização do trabalho; o Capítulo 3 discorre sobre o atual estado de utilização do sistema SCADA para detecção de falhas em aerogeradores bem como o monitoramento de performance, deixando evidente as lacunas na área e como este trabalho encaixa-se neste cenário; a metodologia e modelos empregados são apresentados no Capítulo 4; toda a metodologia proposta é por fim aplicada a dados reais de aerogeradores, com os resultados expostos no Capítulo 5; as conclusões e sugestões para trabalhos futuros são feitas no Capítulo 6.

2 Conceitos Preliminares

2.1 O Aerogerador

Aerogeradores são equipamentos capazes de extrair a energia do vento, convertendo energia cinética em energia elétrica. Em se tratando de monitoramento de performance, é imprescindível conhecer o quanto de energia é possível extrair do vento, o que é possível através da Equação 1 (BURTON et al., 2001). Nela, P é a potência gerada, ξ o rendimento eletromecânico do aerogerador, A é a área varrida pelo rotor, denominada disco atuador, C_p é o coeficiente de potência, ρ a densidade do ar, V_0 a velocidade do vento não perturbado.

$$P = \xi \frac{1}{2} C_p \rho A V_0^3 \tag{1}$$

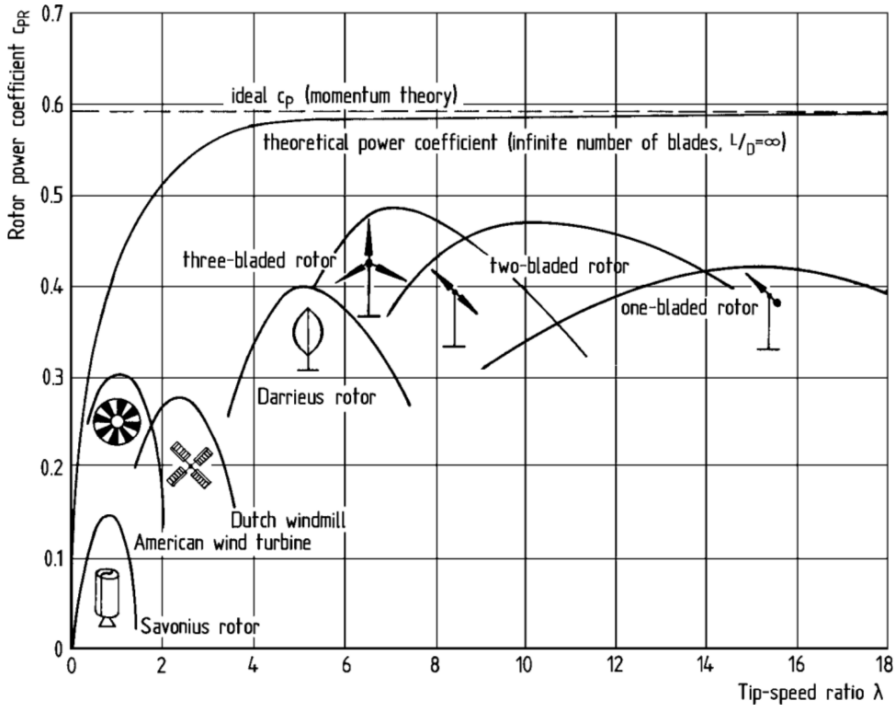
É interessante notar que o valor de C_p é limitado de acordo com a teoria do momento axial para uma turbina eólica ideal. Este limite é dado por $C_p = \frac{16}{27} \approx 59,26\%$ denominado limite de Betz. Uma das hipóteses consideradas no cálculo de Betz é a de que o vento não sofre rotação ao passar pelo rotor, o que não acontece na prática. Considerando isto, Glauert em 1935 recalculou os valores de C_p máximos com base na relação de velocidades na ponta da pá, λ (HANSEN, 2015).

A variação do C_p com λ para diferentes projetos de turbinas é apresentada na Figura 2. Pode-se observar que rotores com 3 pás possuem um C_p de aproximadamente 0.48 quando λ está próximo a 7, superior aos demais modelos e o mais próximo do limite teórico possível. Isto é uma das razões para que turbinas modernas possuam três pás. Outras razões incluem, por exemplo, a redução do peso da caixa de engrenagens, uma vez que esta configuração resulta em menores esforços mecânicos (RAGHEB; RAGHEB, 2011).

Aerogeradores modernos utilizados em aplicações comerciais possuem em sua maioria eixo horizontal, três pás e um sistema de transmissão baseado em caixa multiplicadora. Um esquema com os principais componentes de um aerogerador está na Figura 3. Eles são:

- **Rotor:** O rotor é a denominação dada ao conjunto do hub e pás. Ele é a excitatriz do aerogerador, responsável por transferir a energia do vento através do torque para o eixo principal.
- **Nacele:** Estrutura sobre a torre na qual é montada o sistema de captação, transmissão e

Figura 2 – Coeficiente de potência C_p em função da velocidade de ponta do rotor λ para diferentes modelos de turbinas eólicas.



Fonte – Hau (2005)

conversão de energia, com o objetivo de proteger os componentes das variáveis ambientais.

- **Yaw:** Sistema responsável pela rotação da nacele (controle de giro), especialmente para o posicionamento do rotor de forma perpendicular à direção de incidência do vento. Uma configuração comum para o sistema de yaw é através da instalação de um rolamento esférico sobre o qual a nacele é montada.

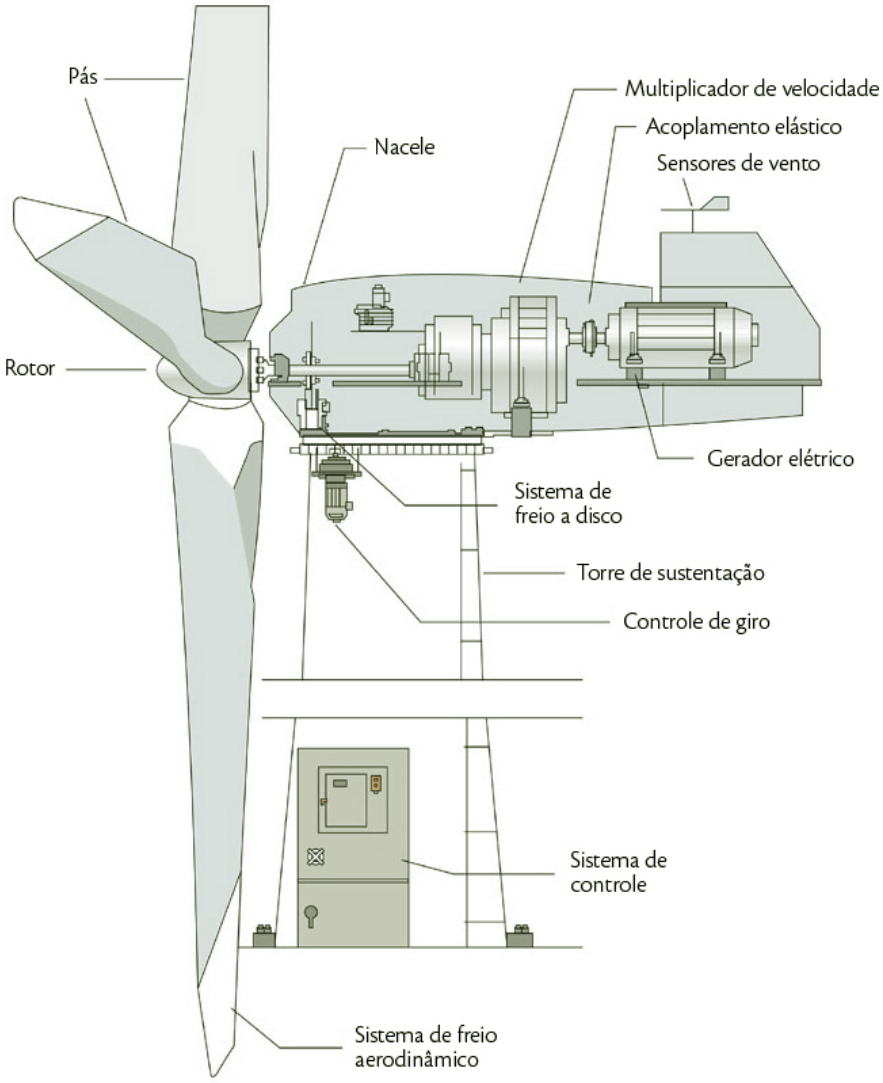
- **Torre:** Estrutura de suporte que permite a captação de vento a alturas elevadas. Podem ser construídas de diversos materiais, como concreto e aço, e diferentes arquiteturas, dentre elas cônicas ou treliçadas, sendo os modelos cônicos tubulares feitos a partir de chapa de aço soldadas os mais comuns.

- **Sistema de Transmissão:** Fazem parte do sistema de transmissão, o eixo principal e a caixa multiplicadora. O eixo principal transmite o torque gerado a partir da rotação das pás em uma frequência baixa. A caixa multiplicadora é responsável por elevar esta frequência de rotação até a entrada do gerador. Há modelos denominados *direct drive* nos quais o torque do rotor é levado diretamente ao gerador, sem a necessidade da caixa multiplicadora.

- **Sistema Elétrico:** Compreende toda a parte de conversão, transmissão de energia e

controle. O principal subcomponente deste sistema é o gerador, que pode ser, por exemplo, assíncrono (gaiola de esquilo).

Figura 3 – Componentes de um aerogerador.



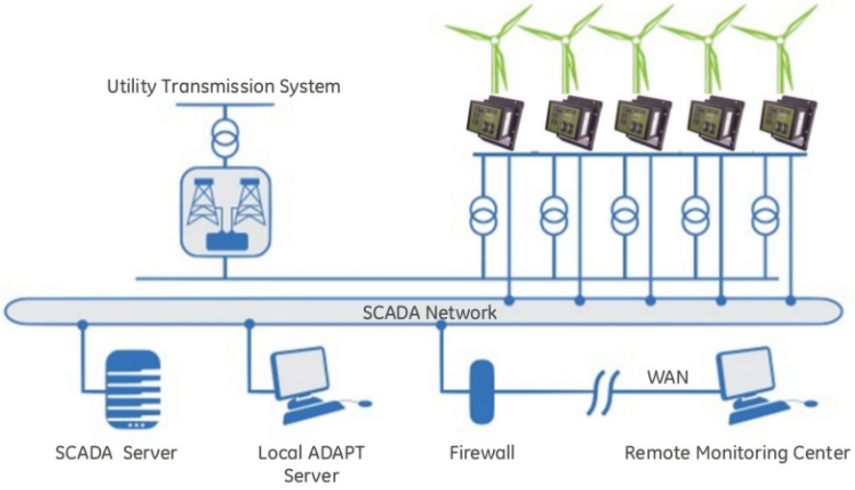
Fonte – <http://www.engenhariacompartilhada.com.br/secoes.aspx?capitulo=30490>. Acesso em 30/08/2016.

2.2 Sistemas SCADA

> A supervisão de processos técnicos foca em mostrar o estado presente, indicando estados permitidos ou não, e tomando medidas apropriadas para evitar danos ou acidentes. Os desvios do comportamento normal dos processos resultam de erros, que podem ser atribuídos a diversas causas. Eles podem resultar em curtos ou longos períodos de mau funcionamento ou até falhas se contramedidas não forem tomadas (ISERMANN, 2011).

Sistemas SCADA são um caso particular de sistemas industriais de controle com o foco em ativos geograficamente dispersos, porém com aquisição e armazenamento centralizado de dados (STOUFFER et al., 2006). Eles permitem que operadores sejam capazes de controlar e armazenar dados de aerogeradores, subestações elétricas e torres anemométricas a partir de um centro remoto, assim como mostrado numa possível arquitetura na Figura 4. Dentre os dados armazenados pelo sistema SCADA, são encontradas variáveis ambientais, como velocidade e direção do vento, valores de produção, temperaturas, pressões, configurações e diversos indicadores de estado.

Figura 4 – Arquitetura de um sistema SCADA



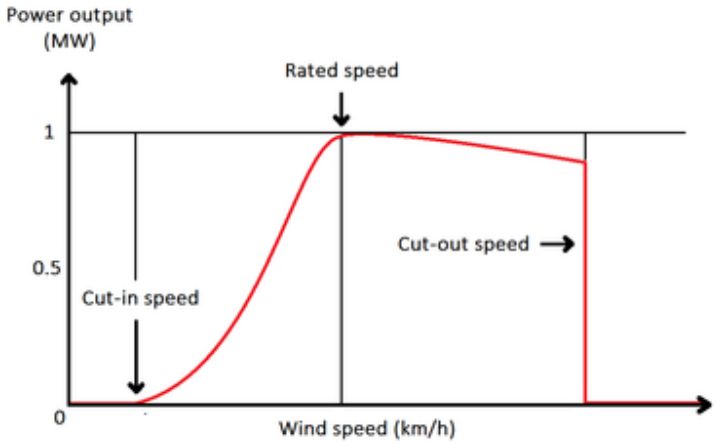
Fonte – <http://windworkshops.sandia.gov/agenda/2013-wind-plant-reliability-workshop-presentations/>. Acesso em 15/04/2016

Com o objetivo de reduzir os custos de produção, sistemas SCADA estão emergindo como uma alternativa a outros sistemas de *condition monitoring* (CMS). Isto porque eles já vêm embarcados nos aerogeradores, não necessitando de um aporte financeiro para a instalação de novos sensores e são capazes de fornecer todo o histórico operacional dos aerogeradores (YANG; JIANG, 2011). Além disso, aspectos que inviabilizavam sua utilização como a baixa capacidade de armazenamento e processamento de dados vêm sendo solucionados através da adoção de novas tecnologias como a nuvem e "big data" (VIHAROS et al., 2013) respectivamente.

2.3 Curvas de Potência

Segundo a IEC (2005) as características de performance de um aerogerador são determinadas por sua curva de potência (CP) e estimativa anual de produção (EAP). A curva de potência apresenta a relação em regime permanente entre a potência elétrica produzida pelo aerogerador e a velocidade de vento incidente, como apresentado pela Figura 5. Para máquinas *onshore* atuais,

Figura 5 – Curva de potência



Fonte – Adaptada de <http://energyeducation.ca/encyclopedia/Wind_power>. Acesso em 30/08/2016.

a norma IEC estabelece que os dados de vento e potência sejam integrados em médias de 10 minutos.

Por permitir avaliar técnico e economicamente (EAP) a capacidade de produção de uma central eólica, a CP é crucial em sua fase de projeto. Ela também é utilizada extensivamente na fase de operação da central como instrumento para monitoramento de performance. As normas IEC-61400-12-1 e IEC-61400-12-2 regem como a CP deve ser caracterizada a partir das torres anemométricas e dos anemômetros sônicos normalmente instalados na nacele dos aerogeradores respectivamente. Neste trabalho, apenas a metodologia de elaboração da CP, denominada método dos bins (IEC, 2005), será considerada. A correção dos valores de potência ativa pela densidade local também sugerida na metodologia da norma IEC não pôde ser realizada em virtude da falta de dados da pressão atmosférica local para o período de análise.

2.4 Modelos de Manutenção

A agência internacional de energias renováveis IRENA[1] estima que o custo para a realização da operação e manutenção dos parques eólicos (O&M) vai de 20% a 25% do LCOE (IRENA, 2014). Uma das razões para tanto está justamente na falta de dados que indiquem a necessidade de manutenção ou reparo do maquinário. O planejamento da manutenção ainda é baseado em cálculos de tendência ou em falhas reais (MOBLEY, 2002).

De acordo com MOBLEY (2002), são três os métodos de gerenciamento da manutenção:

- **Manutenção corretiva:** É uma metodologia reativa, que equivale a não existência de

[1] International Renewable Energy Agency

planejamento de manutenção. A manutenção ocorre apenas após a falha do equipamento, o que a torna a opção mais custosa;

- **Manutenção preventiva:** Todas as manutenções são realizadas com base no número de horas operacionais da máquina, em datas pré-estabelecidas;
- **Manutenção preditiva:** É um sistema de manutenção baseado na condição do equipamento, que se utiliza de meios de monitoramento de condição e técnicas de previsão, acarretando em custos menores.

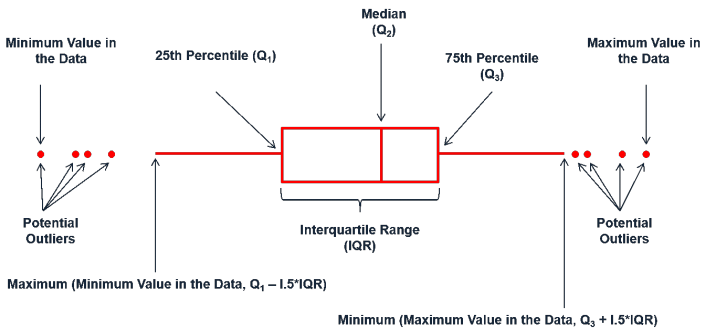
Centrais eólicas aplicam rotineiramente as metodologias de manutenção corretiva e preventiva e em menor grau a manutenção preditiva. Há uma necessidade de transição de paradigma para que técnicas de monitoramento de condição através de vibração mecânica, análise de óleo ou termografia tornem-se rotineiras. O monitoramento de performance via SCADA pode ser de grande valia para a otimização da manutenção preventiva e preditiva, complementando o escopo de monitoramento de outros sistemas de monitoramento por condição (CMS's).

2.5 Diagrama de Caixa

O diagrama de caixa (*box-whisker* ou ainda *box-plot*) é uma representação gráfica que permite a fácil visualização de aspectos da distribuição dos dados, tais como a posição, variabilidade, assimetria e até mesmo valores atípicos (*outliers*) (MAGALHÃES, 2005).

Um *boxplot* é apresentado na Figura 6, juntamente com sua terminologia associada. Para a construção de um *boxplot*, é necessário a definição de um retângulo cujo limite inferior e superior correspondem ao primeiro e ao terceiro quartil, dados por Q_1 e Q_3 respectivamente. A mediana dos dados (Q_2) é dada pelo traço dentro do retângulo.

Figura 6 – Diagrama de caixas



Fonte – <http://bradleyboehmke.github.io/tutorials/descriptives_numeric>. Acesso em 30/08/2016.

Ao intervalo contendo 50% das observações dá-se o nome de intervalo interquartílico, ou IQR, definido por $IQR = Q_1 - Q_3$. É comum a definição de uma região onde os dados são considerados "bem comportados", delimitada pelos bigodes (*whiskers*).

Autores como TUKEY (1977) utilizam além do conceito de *outliers*, o conceito de *outliers* severos, apresentados nas Equações 2a e 2b. Qualquer valor dentro do intervalo dado pela Equação 2a é dito "bem comportado". Já os que excedem os limites dados pelas Equações 2a e 2b são denominados *outliers* e *outliers* severos, respectivamente. É importante ressaltar que nessas equações, os valores que multiplicam o IQR ($1,5$ e $3,0$) são empíricos. Por esta razão, os multiplicadores do IQR podem ser substituídos por um valor genérico α, definido de acordo com a característica do problema estudado.

$$]Q_1 - 1,5 \times IQR, Q_3 + 1,5 \times IQR[\tag{2a}$$

$$]Q_1 - 3,0 \times IQR, Q_3 + 3,0 \times IQR[\tag{2b}$$

3 Revisão Bibliográfica

3.1 Monitoramento de Performance de Aerogeradores

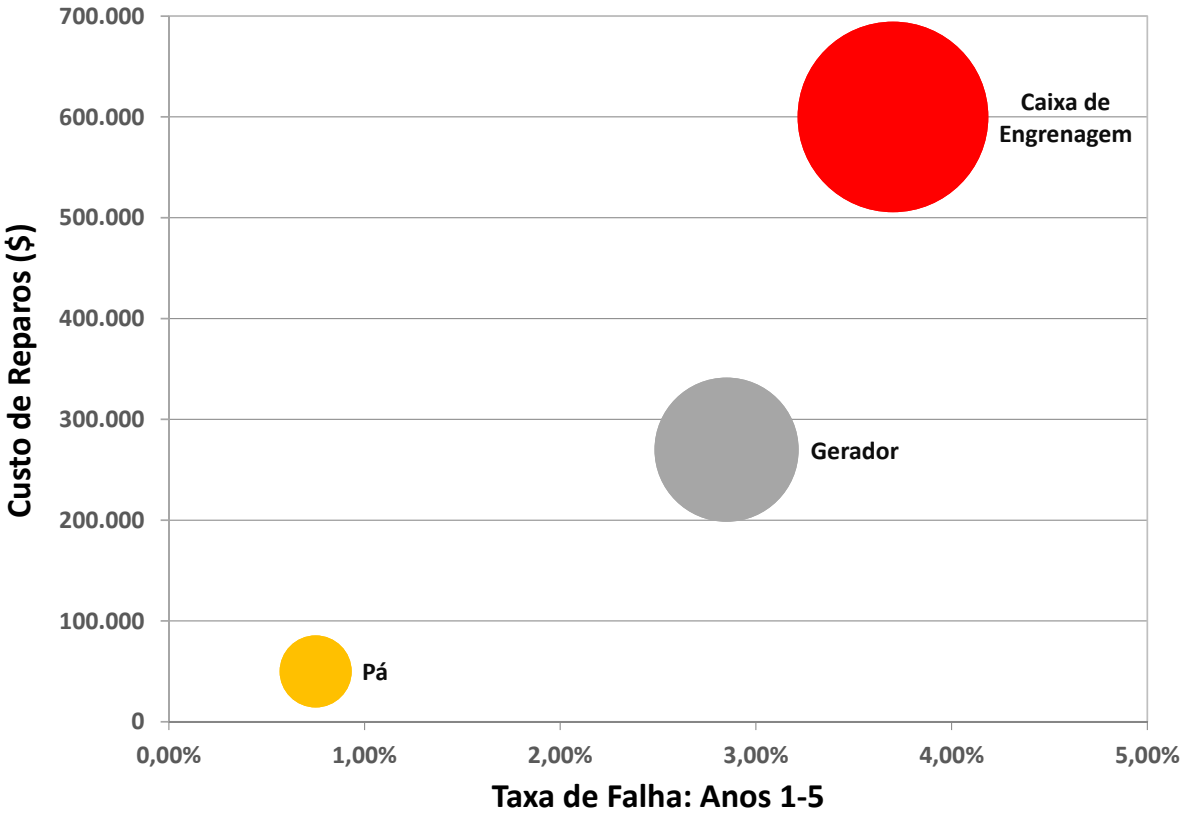
A última década foi bastante prolífica com relação aos estudos dedicados à aplicação do sistema SCADA com fins de monitoramento de performance e detecção de falhas em aerogeradores. Hyers et al. (2006) mostraram como o setor eólico vem importando técnicas de monitoramento de condição e diagnóstico com sucesso de outras indústrias, a exemplo da aeronáutica. Porém, as menções ao sistema SCADA resumiam-se apenas à sua capacidade de aquisição de dados. Desde então, a utilização do sistema SCADA expandiu-se do controle em nível de processo, passando a englobar o monitoramento e diagnóstico de falhas. Hoje, o campo é tido como uma maneira viável de redução do LCOE (YANG; COURT; JIANG, 2013).

Uma das maneiras de se reduzir o LCOE de um empreendimento eólico é através da redução de custos operacionais. Segundo a *Det Norske Veritas and Germanischer Lloyd* (DNV-GL), a taxa de falhas dos componentes principais do aerogerador nos seus primeiros cinco anos de vida útil pode chegar a 3,5% para caixas de engrenagens, com custos médios na ordem dos USD 600.000, conforme Figura 7 (GRAVES, 2014). Tendo em vista mitigar esses riscos, estudos utilizando dados do sistema SCADA com foco no monitoramento de caixas de engrenagem (WANG; INFIELD, 2013; WILKINSON et al., 2014; LETZGUS, 2015), geradores (GUO; INFIELD; YANG, 2012; SCHLECHTINGEN; SANTOS, 2011), pás (YANG; COURT; JIANG, 2013) e rolamentos (BUTLER; RINGWOOD; O'CONNOR, 2013; ZHANG; WANG, 2014) vêm sendo conduzidos.

Outra forma de minimizar o LCOE de um projeto eólico é com o aumento de produção energética. Uma vez que a modificação da capacidade instalada ou dos atributos do aerogerador como aumento do diâmetro do rotor e comprimento da torre são inviáveis em empreendimentos já construídos, resta como alternativa evitar perdas de produção. Através do monitoramento de performance, é possível avaliar o quão eficiente o aerogerador opera em relação a valores ideais ou contratuais de desempenho. Uma das ferramentas mais utilizadas com esta finalidade é a curva de potência (CP).

Lydia et al. (2014) realizaram uma revisão sobre o estado da arte em curvas de potência e as classificaram tanto quanto à sua finalidade para cálculo energético, monitoramento de performance ou previsão de potência, quanto à natureza dos métodos utilizados em seu cálculo, que podem ser paramétricos ou não conforme apresentado na Figura 8. Em sua essência, grande

Figura 7 – Custo total versus taxa de falha de componentes principais nos primeiros 5 anos de operação de parques eólicos da DNV-GL



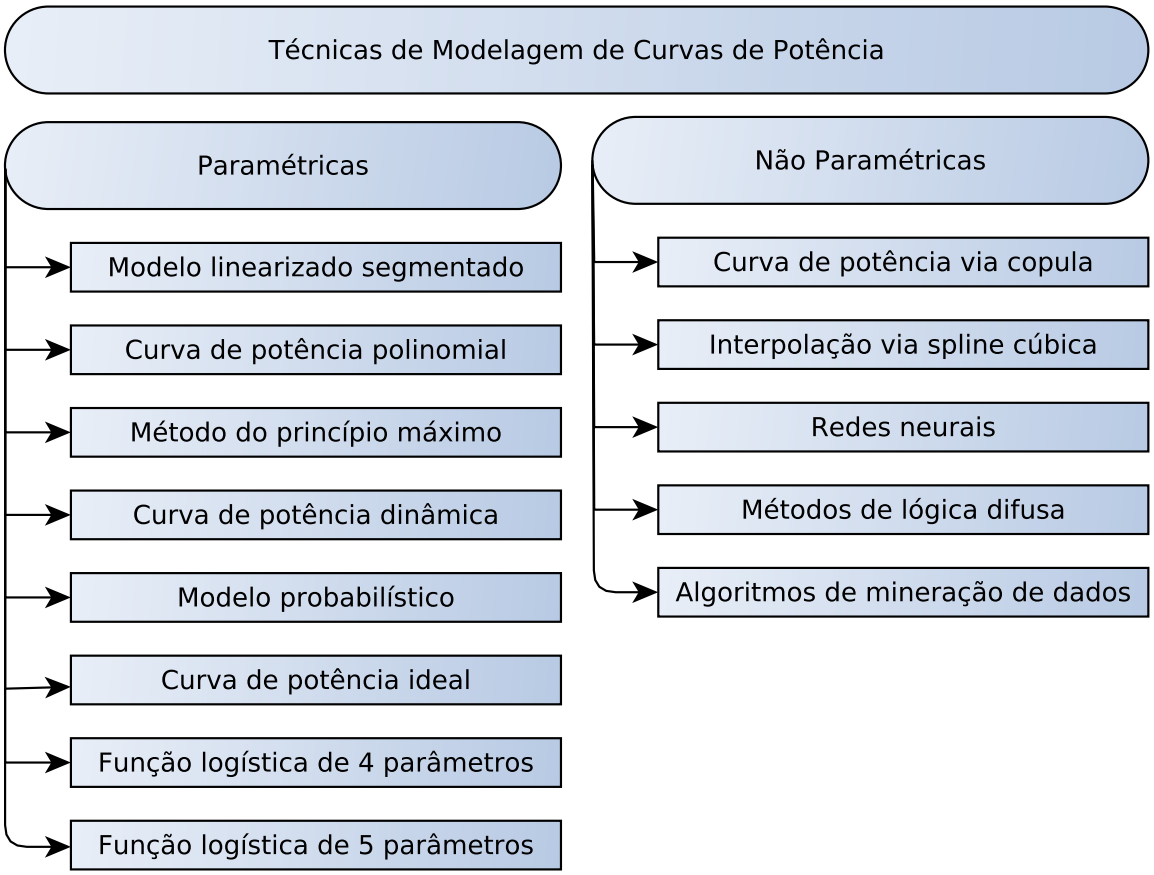
Fonte – Adaptado de Graves (2014)

parte dos sistemas de monitoramento de performance ou detecção de falhas baseados em curvas de potência operam estabelecendo uma curva referencial ou modelo do modo de operação normal (LETZGUS, 2015).

Aplicando o conceito de modelo normal, Kusiak e Verma (2013) propuseram a elaboração não só de curvas de potência para o monitoramento de aerogeradores, mas também de pitch e torque através da aplicação do k-means clustering, utilizando como métrica para detecção de *outliers* a distância de Mahalanobis. Os gráficos das distribuições bimodais de cada aerogerador foram então resumidos e comparados com base nos seus terceiro e quarto momentos de inércia, a assimetria e a curtose. Pontos significativamente mais afastados do local de maior concentração indicam alterações na curva de monitoramento e, portanto, a possibilidade de existência de falha. O monitoramento da performance através do tempo é realizado a partir de um gráfico de controle Hotelling T^2. Uma das grandes vantagens do método proposto é a sistematização da análise, que pode ser realizada sem a necessidade de verificação gráfica das curvas de monitoramento.

Uluyol et al. (2011) utilizaram um ajuste polinomial para a geração de resíduo a partir dos dados observacionais de potência dos aerogeradores estudados nos casos em que a curva

Figura 8 – Métodos para elaboração de curva de potência.



Fonte – Adaptado de Lydia et al. (2014)

do fabricante não estava disponível (como em situações de remodelação, troca massiva de componentes ou modificações na estratégia de controle). O método consiste em estabelecer um modelo base de resíduo cujos limites operacionais são definidos a partir de cada bin de velocidade e determinados por um múltiplo do desvio padrão, neste caso 3σ. A partir do cálculo de estatísticos como média, desvio, assimetria ou curtose, é possível gerar indicadores de condição. Esta metodologia permitiu a identificação de uma falha no eixo axial de uma gearbox com 20 dias de antecedência. Porém, devido à sua natureza ampla, o método é incapaz de associar a anomalia à uma falha particular.

Através da construção de uma curva de potência utilizando regressão gaussiana, Butler, Ringwood e O'Connor (2013) geraram uma série de resíduos utilizada para a detecção de falhas, assim como Uluyol et al. Como indicador de condição, foi utilizada a soma acumulada do resíduo, que deve oscilar em torno de zero para cenários de funcionamento normal e tender a valores negativos para situações de falha. Isto permitiu a detecção de uma anomalia no rolamento principal do aerogerador meses antes da falha ocorrer.

Em 2007, Singh, Bhatti e Kothari alertaram para o potencial de utilização redes neurais para tarefas de detecção e diagnóstico de falhas em aerogeradores e propuseram uma arquitetura de rede *multi layer perceptron* (MLP) com um algoritmo de treinamento utilizando um filtro de Kalman extendido. Marvuglia e Messineo (2012) expandiram o estudo aplicando, além da arquitetura de rede MLP, a *generalized mapping regressor* (GMR) e a *general regression neural network* (GRNN). Além disso, em seu trabalho, eles propuseram a elaboração de uma única curva de potência para monitoramento on-line de todo o parque, com o auxílio de um gráfico de controle para o resíduo entre as estimativas obtidas pelos modelos e os dados observacionais de potência. Letzgus (2015) gerou modelos de potência a partir de redes com arquiteturas *nonlinear autoregressive exogenous* (NARX), obtendo resultados satisfatórios tanto para detecção de ineficiências aerodinâmicas geradas por gelo depositado sobre as pás da turbina, quanto falhas na caixa de engrenagem.

Um aspecto em comum entre os modelos não paramétricos propostos é a utilização de mais de uma variável regressora. Isto impõe desafios na seleção e garantia de qualidade dos dados, pois se deseja que na construção de modelos normais apenas dados sem falha sejam utilizados. Nesse sentido, Karlsson (2015) desenvolveu um filtro automático baseado em técnicas de *clustering* e distância da Mahanolabis capaz de remover pontos cujos padrões divergem dos dados normais para até 8 dimensões. Yampikulsakul et al. (2014) removeram o efeito dos outliers em seu modelo estatístico não paramétrico da potência através da utilização de um vetor de regressão de apoio baseado nos mínimos quadrados. A definição dos limites entre os status de condição normal ou não do aerogerador é feita após o exame da variação das condições operacionais sensíveis a alterações nas variáveis ambientes externas, as mesmas utilizadas para regressão da potência. Estes limites foram empregados na construção de gráficos de controle utilizados para monitoramento on-line de aerogeradores.

Naturalmente, sistemas completos de monitoramento de condição também foram propostos visando eliminar tanto os custos de oportunidade de geração quanto as despesas geradas pelo funcionamento ineficiente dos aerogeradores. Zaher et al. (2009) propuseram a utilização de um sistema multiagente (MAS), onde a potência ativa e cada componente principal do aerogerador correspondem a um agente. Estes agentes colaboram entre si para gerar apenas uma saída indicando a existência ou não de falha ao operador, de forma robusta.

3.2 Enunciado do Problema

Mesmo com todas as aplicações bem sucedidas e avanços na análise de dados SCADA, existem muitos desafios em aberto que impõem dificuldades adicionais ao seu emprego como ferramenta de monitoramento tais como (WANG; SHARMA; ZHANG, 2014):

- As arquiteturas de sistemas SCADA podem diferir de aerogerador para aerogerador;
- Os aerogeradores estão sujeitos às condições de operação;
- Os dados observacionais são geralmente armazenados com um intervalo de integração de 10 minutos, perdendo informações de alta frequência;
- Os aerogeradores apresentam alta taxa de falsos alarmes;
- Não existe um consenso de técnicas para análise de dados SCADA.

Além disso, do ponto de vista do monitoramento de performance é desejável compreender as origens da variação de eficiência com que o aerogerador é capaz de extrair energia do meio. Deseja-se que anomalias na performance possam ser associadas a falhas específicas, bem como descartar falsos alarmes gerados pela variabilidade inerente ao vento ou outros fatores ambientais.

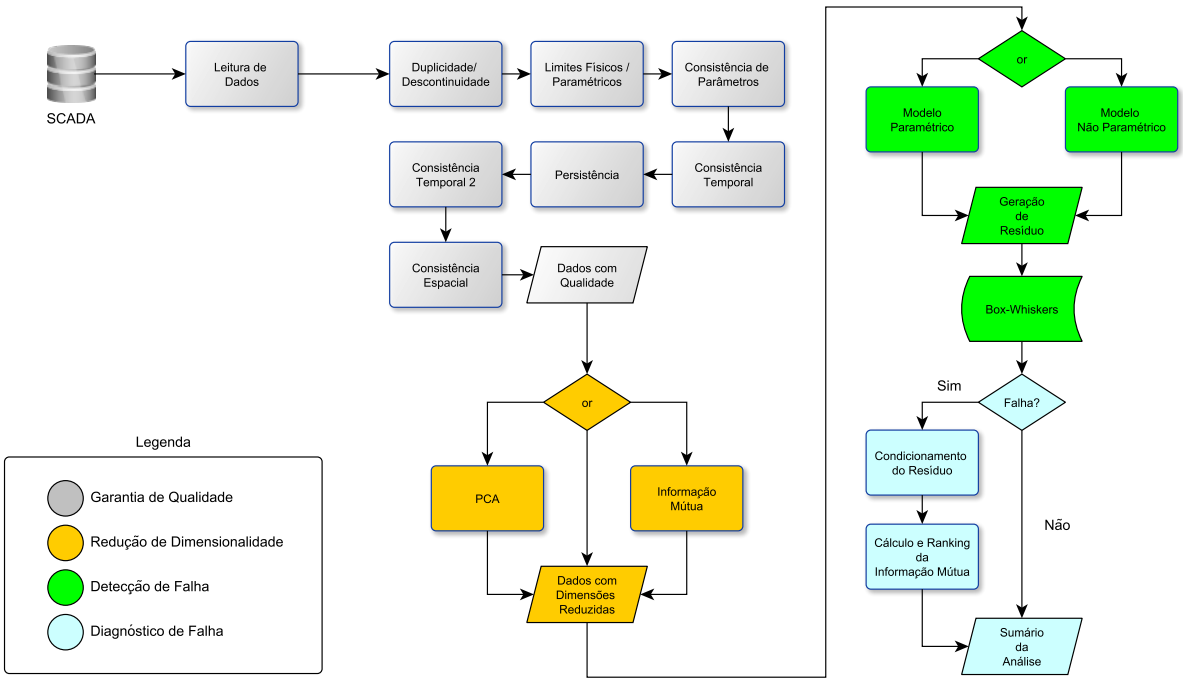
Nesse sentido, este trabalho visa não só detectar falhas na performance de aerogeradores, mas também estabelecer relações de causa e consequência entre a falha e sua causa raiz. Como o volume de dados empregados para tal finalidade é massivo, o problema será abordado de forma automática e objetiva.

4 Metodologia e Modelos

4.1 Metodologia

São apresentados, nesta seção, os procedimentos adotados para elaboração do sistema de monitoramento de performance. Uma visão detalhada das etapas que compreendem o sistema é mostrada na Figura 9. Os procedimentos podem ser divididos em quatro grandes grupos: garantia de qualidade dos dados, redução de dimensionalidade, detecção de falhas e uma nova técnica de diagnóstico de falhas, cada um representada por uma cor distinta e apresentado na Seção 4.2.

Figura 9 – Fluxograma com o caminho percorrido pelos dados SCADA até o resultado do diagnóstico do aerogerador.



Fonte – Próprio autor.

Os dados de entrada são provenientes de aerogeradores reais, multimegawatt, com caixa de engrenagem, gerador de indução com *active stall*[1]. Por razões de confidencialidade, apenas informações suficientemente necessárias serão apresentadas.

[1] Técnica de regulação de potência através da variação do ângulo de pitch.

4.2 Modelos

4.2.1 Garantia de Qualidade

Um dos problemas com relação à elaboração de modelos para monitoramento de condição é a propagação de erros. Isto acontece quando dados anômalos são utilizados para o treinamento do modelo, ocasionando um alto índice de falsos alarmes. Exemplos de possíveis problemas relativos à qualidade dos dados do sistema SCADA são (GRAY et al., 2011):

- Formato inconsistente de dados;
- Outliers estatísticos;
- Limites de plausabilidade excedidos;
- Repetições de um único ou um bloco de dados;
- Frequência de medições variável.

Através de processos de garantia de qualidade é possível identificar e corrigir erros sistemáticos, responsáveis pela introdução de um viés nos valores da série, bem como eliminar erros grosseiros, associados ao mal funcionamento de sensores ou manuseio dos dados. Apesar da importância da aplicação de tal processo, o único consenso sobre o tema é que "*são tarefas tediosas e que consomem tempo*" (JIMÉNEZ et al., 2010).

Utilizando a potência ativa para a determinação de CPs baseadas em condições operacionais, Paiva, Rodrigues e Palma (2014) propuseram três critérios para garantir a qualidade dos dados empregados. O primeiro diz respeito à utilização do limite de Betz para remoção de valores de potência impossíveis de serem atingidos. Já o segundo remete-se à delimitação de bandas para dados *outliers* e *outliers* severos através do diagrama de caixas, cujos limites são dados pelas Equações 2a e 2b (seção 2.5) respectivamente. Por fim, como terceiro critério, foi adotada a exclusão de dados pertencentes ao grupo com menor densidade de pontos em histogramas multimodais. Este último critério foi empregado ao se constatar falhas no emprego de filtros estatísticos convencionais, que não são capazes de distinguir entre diferentes modos de operação do aerogerador.

Tratando-se de garantia de qualidade para dados anemométricos, diversos estudos no campo de climatologia e meteorologia já foram realizados (VEJEN et al., 2002; JIMÉNEZ et al., 2010). O trabalho de Moraes (2014) destaca-se por propor uma metodologia objetiva de garantia de qualidade para dados anemométricos voltada para o cálculo de geração de energia. Ela consiste na verificação de:

- Lacunas na série temporal;
- Limites físicos e paramétricos;
- Consistência de parâmetros;
- Consistência temporal de curtíssimo prazo;
- Persistência;
- Consistência temporal de curto prazo;
- Consistência espacial;

Neste trabalho, foi empregada uma generalização do trabalho de Moraes (2014) para que não somente os dados anemométricos sejam submetidos a um processo de garantia de qualidade, mas também dados de potência ativa do aerogerador, incorporando o primeiro e uma variante do segundo critério introduzido por Paiva, Rodrigues e Palma (2014). Para tanto, o cálculo do valor α foi substituído por um método robusto de determinação das bandas de limites do *box-whiskers* assimétricas (HUBERT; VANDERVIEREN, 2008). É interessante notar que o método permite determinar os limites inclusive para variáveis com assimetria elevada, e que eles recaem no critério de Tukey (TUKEY, 1977) para distribuições simétricas (dado pelas Equações 2). Os casos extremos em que o IQR apresentou-se nulo foram contornados através da utilização da mediana da série no lugar do IQR. Os novos limites são dados pelas Equações 3:

$$]Q_1 - 3,0e^{-4MC} \times IQR, Q_3 + 3,0e^{3MC} \times IQR[\quad \text{Se } MC \geq 0 \tag{3a}$$

$$]Q_1 - 3,0e^{-3MC} \times IQR, Q_3 + 3,0e^{4MC} \times IQR[\quad \text{Se } MC < 0 \tag{3b}$$

Onde MC representa a *medcouple*, uma medida de assimetria da distribuição de uma variável (HUBERT; VANDERVIEREN, 2008). O seu cálculo foi realizado através da utilização da *toolbox* LIBRA (VERBOVEN; HUBERT, 2004). Notar que para $MC = 0$ obtém-se a Equação 2b para outliers extremos, sugerida por Paiva, Rodrigues e Palma (2014).

4.2.2 Redução de Dimensionalidade

Dimensão é uma palavra com múltiplos significados e pode apresentar diversas facetas a depender do contexto em que é empregada. Empiricamente, sua definição está associada à ideia de extensão, tamanho ou grandeza. No contexto da estatística ou do aprendizado de máquinas, a dimensionalidade de um conjunto de dados é definida pelo número de variáveis medidas de interesse.

Com os avanços na capacidade de armazenamento de dados, ocorre um fenômeno denominado de sobrecarga de informação. Isso significa que o número de variáveis à disposição excede aquele necessário para se entender o fenômeno estudado (FODOR, 2002). Não obstante a sobrecarga de informação, há ainda a sobrecarga de algoritmos comumente utilizados para a mineração de dados. A alta dimensionalidade reduz a eficiência da maioria dos algoritmos, que são por natureza caros computacionalmente (MAIMON; ROKACH, 2010). Torna-se imperativo, portanto, no contexto do diagnóstico de falhas, que haja uma redução do número de variáveis selecionadas do conjunto de dados do SCADA para a regressão de modelos. Para tanto, foram selecionadas duas técnicas: a análise de componentes principais (PCA[2]) e a informação mútua, apresentadas a seguir.

4.2.2.1 Análise de Componentes Principais

A análise de componentes principais, ou simplesmente PCA, é uma técnica não paramétrica bastante difundida na redução de dimensionalidade por ser de fácil aplicação e capaz de lidar com dados altamente correlacionados. Ela consiste em encontrar uma nova base através de uma transformação linear que seja capaz de explicar a maior parte da variância do conjunto original de dados (DING, 2014).

De acordo com Shlens (2014), dada a matriz de dados $\mathbf{X}_{m\times n}$ com m medições/amostras e n variáveis, a PCA consiste em encontrar a matriz ortogonal de componentes principais $\mathbf{P}$, de forma que $\mathbf{Y}$, dado por $\mathbf{Y} = \mathbf{PX}$, possua matriz de covariância $\mathbf{C_Y} = \frac{1}{m}\mathbf{YY^T}$ diagonal. Por definição,

$$\mathbf{C_Y} = \frac{1}{m}\mathbf{YY^T} \tag{4a}$$

$$= \frac{1}{m}(\mathbf{PX})(\mathbf{PX})^\mathbf{T} \tag{4b}$$

$$= \mathbf{PC_X P^T} \tag{4c}$$

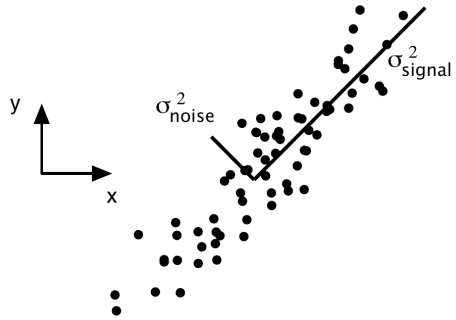
Decompondo a matriz de covariância $\mathbf{C_X}$ através do teorema espectral:

$$\mathbf{C_X} = \mathbf{E\Lambda E^T} \tag{5}$$

Conclui-se portanto que $\mathbf{P} \equiv \mathbf{E^T}$. Logo, $\mathbf{P}$ é uma matriz ortogonal dos autovetores da matriz de covariância, enquanto que Λ é a matriz diagonal com seus autovalores ordenados, isto

[2] *Principal Component Analysis*

Figura 10 – Exemplo de sinal bidimensional com presença de ruído.



Fonte – (SHLENS, 2014, p.4)

é, $\lambda_1 \leq \ldots \leq \lambda_n$. Esta característica da decomposição espectral permite interpretar cada autovalor como a variância associada ao seu autovetor correspondente. Desta forma, a Equação 6 representa a proporção cumulativa da variância explicada pelas primeiras k componentes principais, onde a variância total é dada pelo traço da matriz de covariância. Neste trabalho, uma proporção cumulativa de 95% foi adotada como critério de retenção das primeiras k variáveis.

$$\sum_{i=1}^{k} \frac{\lambda_i}{\mathrm{Tr}(\mathbf{C_X})} \tag{6}$$

Uma análise atenta da Equação 6 mostra ainda o *trade-off* existente na redução de dimensionalidade por trás da PCA. É abdicando das componentes principais que menos explicam a variância dos sinais, que a dimensionalidade é reduzida sem perda significativa de informação. Isto é facilmente percebido ao se avaliar a Figura 10, onde a maior variância está sob a linha de melhor ajuste indicando que $y = f(x)$ e $x = f^{-1}(y)$, ou seja, que apenas uma variável é necessária para descrever o fenômeno em questão contanto que a variância do ruído (σ^2_{noise}) seja desprezível quando comparada à variância do sinal (σ^2_{signal}).

Apesar de todos os benefícios trazidos pela PCA, alguns cuidados devem ser tomados em sua aplicação. Como a variância de um sinal é dependente de sua escala, é necessária a estandardização do sinal antes da aplicação da PCA, de forma que cada variável possua média zero e desvio padrão um. Assim, todas as variáveis tornam-se equiparáveis. Além disso, o fato de haver uma mudança de base faz com que as componentes principais percam qualquer caráter físico, dificultando a interpretação dos resultados (FODOR, 2002).

4.2.2.2 Informação Mútua

A informação mútua (IM) é um conceito da teoria da informação que representa a quantidade de informação que uma variável aleatória contém sobre outra. Mais especificamente, dadas duas variáveis aleatórias X e Y, a informação mútua $I(X;Y)$ é a entropia relativa entre a distribuição conjunta $p(x,y)$ e o produto das distribuições marginais, dado por $p(x)p(y)$ (COVER; THOMAS, 1991).

$$I(X;Y) = -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2\frac{p(x,y)}{p(x)p(y)} \tag{7a}$$

$$= D(p(x,y)\|p(x)p(y)) \tag{7b}$$

$$= H(X) - H(X\|Y) \tag{7c}$$

A entropia relativa $D(p\|q)$, ou distância de *Kullback-Leibler*, mede a distância entre duas funções de densidades de probabilidade. Ela é uma medida da ineficiência de se assumir uma distribuição q enquanto a real distribuição é p, tal como expresso na Equação 8.

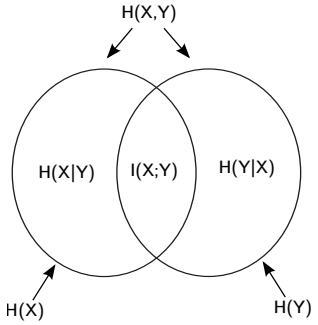
$$D(p\|q) = \sum_{x\in X} p(x)\log_2\frac{p(x)}{q(x)} \tag{8}$$

Note que a informação mútua também pode ser definida em função da entropia $H(X)$ e da entropia condicional $H(X\|Y)$. A entropia pode ser descrita como o "auto conhecimento" que uma variável possui sobre si mesma. Matematicamente, ela é dada pela Equação 9, função da densidade de probabilidade da variável aleatória X. A relação entre a diferentes medidas de entropia e a informação mútua é apresentada intuitivamente na Figura 11. É importante ressaltar a introdução do conceito de entropia conjunta $H(X,Y)$, medida de incerteza associada a duas variáveis aleatórias (área ocupada pelos discos no diagrama de Venn da Figura 11).

$$H(X) = -\sum_{x\in X} p(x)\log_2 p(x) \tag{9}$$

Diversas técnicas já foram propostas empregando a informação mútua para a redução de dimensionalidade (YANG; MOODY, 1999; PENG; LONG; DING, 2005; BENNASAR; HICKS; SETCHI, 2015). Foi selecionado, porém, o critério de maximização da informação mútua entre as variáveis do SCADA e a potência ativa. Isto porque, se a variável do SCADA for independente com relação à potência, a informação mútua é zero, não trazendo informação alguma. O oposto também é verdadeiro e à medida que existam relações funcionais mais estreitas

Figura 11 – Relação entre a entropia, entropia condicional, entropia conjunta e informação mútua de duas variáveis aleatórias.



Fonte – Adaptado de Cover e Thomas (1991)

entre as variáveis do SCADA e a potência ativa, a IM aumenta, indicando a relevância da variável para a modelagem de curvas de potência.

4.2.3 Detecção de Falhas

De posse de dados com qualidade assegurada e de um número reduzido de dimensões para a análise da eficiência de máquina, é elaborado um modelo para detecção automática de falhas. Ele compreende três etapas:

1. Modelagem de curva de potência;
2. Geração de resíduo;
3. Classificação do resíduo.

Inicialmente, foi estabelecida uma relação entre a potência ativa e suas variáveis regressoras, através da modelagem da curva de potência. Foram considerados dois tipos de abordagens para a modelagem das CPs: paramétricas e não paramétricas. Dentre os modelos paramétricos adotados estão o linear segmentado e o ajuste polinomial de quinto e nono grau, enquanto o modelo não linear empregado foi o de redes neurais. Além dos modelos citados, a CP fornecida pelo fabricante do aerogerador será utilizada como referência na comparação da eficácia dos modelos propostos. Os modelos de CP utilizados bem como os dados de entrada destes modelos estão resumidos na Tabela 1.

Em seguida é necessário comparar o resultado do melhor modelo de curva de potência obtido com os dados observacionais. Isto é feito a partir do cálculo do resíduo. Tomando proveito da metodologia objetiva de garantia de qualidade de dados, a mesma técnica do *box-whiskers* é

Tabela 1 – Modelos de CP e dados de entrada considerados no trabalho.

	Dados de Entrada		
Modelo de CP	PCA	IM	Velocidade do vento
LS	X	X	
NN	X	X	X
P_9 e P_5			X
P_{man}			X

Fonte – Próprio autor.

aplicada para os dados do resíduo, o que possui a vantagem de não necessitar supor qual a sua distribuição, diferentemente da metodologia proposta por Uluyol et al. (2011) onde foi assumido uma distribuição normal do resíduo e bandas simétricas 3σ para delimitar o seu intervalo de controle para cada bin de velocidade. Caso os dados excedam os limites, eles serão tidos como anômalos, prosseguindo para a etapa de diagnóstico abordada na Seção 4.2.4.

4.2.3.1 Regressão Multilinear baseada nos Mínimos Quadrados

A regressão multilinear (MLR) consiste em encontrar o polinômio de grau *um* baseado nas variáveis regressoras $X_1, X_2, \ldots, X_n$, de forma que o erro médio quadrático entre a variável estimada $\hat{Y}$ e os dados observacionais Y seja minimizado (SERGENT et al., 1995).

Matematicamente lê-se : *Encontre*

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \tag{10}$$

com

$$e_i = Y_i - \hat{Y}_i \tag{11}$$

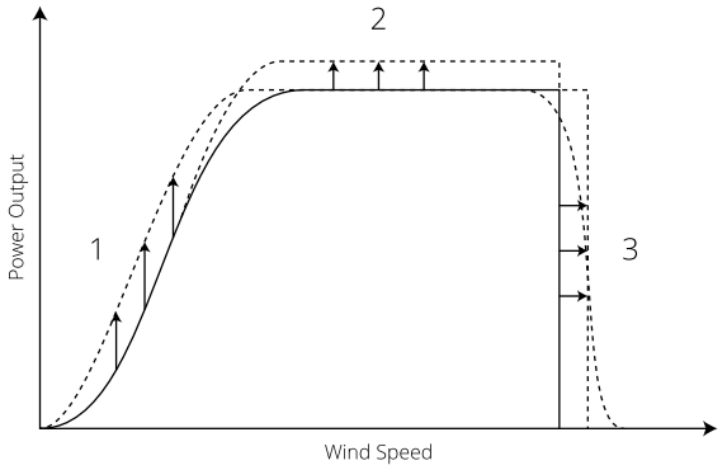
e $i = 1...n$ *de forma que*

$$\begin{cases} \frac{\partial e^2}{\partial \beta_0} = 0 \\ \\ \frac{\partial e^2}{\partial \beta_n} = 0 \end{cases} \tag{12}$$

Para um modelo linear de potência com múltiplas variáveis regressoras a Equação 10 poderia ser utilizada. Devido à forma sigmoidal da CP, esta não seria uma boa aproximação,

tornando a adoção de um modelo segmentado mais eficaz. Desta forma, foi segmentada a curva de potência em múltiplas partes conforme Tabela 2, e para cada parte, calculado um modelo linear distinto. A escolha dos intervalos foi feita de forma a agregar robustez ao modelo e minimizar os erros nas zonas com maior curvatura, bem como privilegiar a zona 1 da Figura 12, região com maior frequencia de incidência de vento e mais propensa à ocorrência de falhas.

Figura 12 – Zonas de falhas distintas de uma curva de potência.



Fonte – Breeze (2015)

Tabela 2 – Quantidade de segmentações por zonas de velocidade para modelo linear segmentado de CP.

Intervalo de Velocidade (m/s)	Número de Segmentações
$0 \leq v \leq v_{cut-in}$	3
$v_{cut-in} < v \leq v_{rated}$	11
$v_{rated} < v \leq v_{cut-out}$	3
$v_{cut-out} < v \leq v_{max}$	2

Fonte – Próprio autor.

Caso as variáveis $X_1, X_2, \ldots, X_n$ da Equação 10 sejam substituídas por $x^1, x^2, \ldots, x^n$, o mesmo método de minimização dos erros pode ser empregado para encontrar os coeficientes de um polinômio de grau n. Neste trabalho, foram selecionados os polinômios de quinto e nono graus, conforme indicado no trabalho de Pelletier, Masson e Tahan (2016).

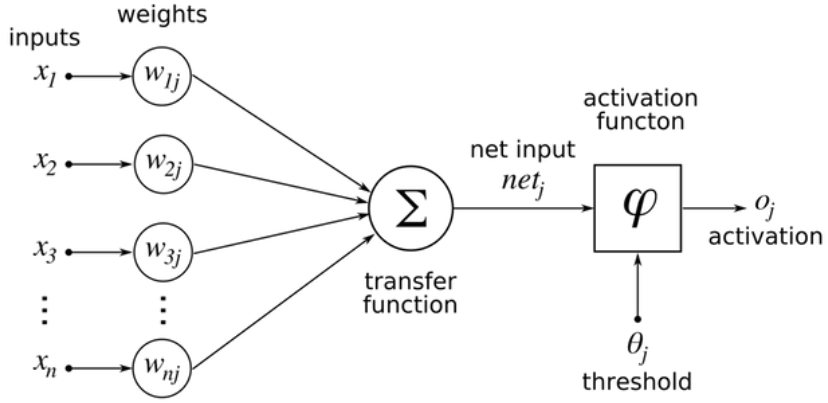
4.2.3.2 Redes Neurais Artificiais

Redes neurais artificiais (NN[3]) são sistemas computacionais de inteligência artificial capazes de simular operações realizadas pelo cérebro humano, tais como reconhecimento, cálculos e tomadas de decisões. Analogamente ao sistema biológico, uma rede neural artificial é composta por:

- Neurônios: são as unidades básicas de processamento e manuseio de informações altamente interligadas;
- Sinapses: impulsos responsáveis pela transmissão de informação entre neurônios.

O neurônio pode ser modelado conforme indicado na Figura 13. É nos dendritos e terminais dos axônios que ocorrem as sinapses. A intensidade das sinapses que ocorrem nos dendritos é representada pelo peso w. O viés θ é um parâmetro que indica um peso e pode assumir qualquer valor inclusive nulo. O resultado do neurônio é conduzido através do axônio e é dado pela variável o, após sofrer a alteração de uma função transferência denominada função de ativação, capaz de codificar relações complexas. Como unidade básica neural, o conjunto de neurônios interligados é denominado rede neural.

Figura 13 – Esquema de um neurônio artificial.



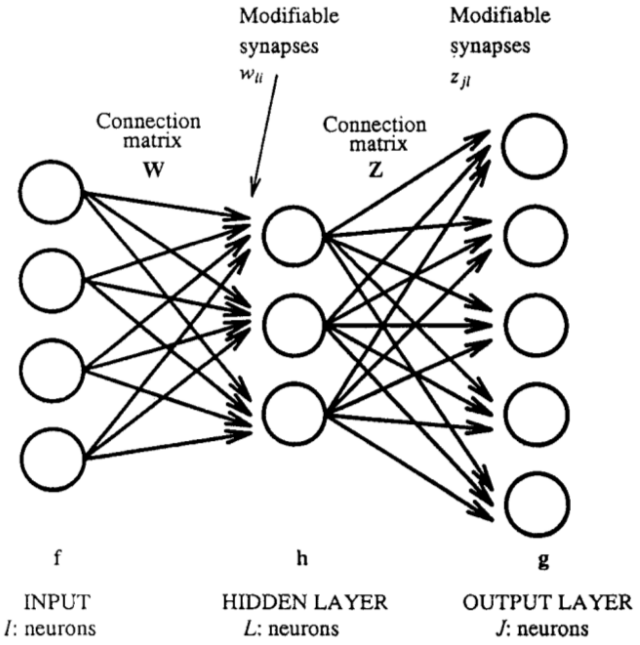
Fonte – <www.innoarchitect.com/artificial-intelligence-deep-learning-neural-networks-explained>. Acesso em 01/09/2016.

Um dos primeiros modelos de redes neurais propostos foi o do *perceptron*, utilizado para simular as capacidades perceptivas da retina. Este modelo é composto de uma camada de neurônios de entrada ou estímulo (retina artificial), e outra de saída. A resposta do perceptron é, portanto, uma função do estímulo captado pela camada de entrada e dos pesos atribuídos às conexões neurais, que determinarão sua capacidade sináptica (ABDI, 1994).

[3]Neural Network

Uma grande limitação do modelo perceptron é sua incapacidade de aproximar relações não lineares. Isto foi superado através da inserção de uma camada intermediária denominada *hidden layer*, conforme apresentado na Figura 14, o que só foi possível após a criação de algoritmos, como o *backpropagation*, para treiná-la. Estes algoritmos são capazes de ajustar os pesos dos neurônios baseados no conjunto de dados de entrada a serem aprendidos. Mais especificamente, o *backpropagation* utiliza um algoritmo para atualizar os pesos e viés da rede de forma retroativa, com base no gradiente da função objetiva. Um método de otimização bastante utilizado para o treino de redes neurais, principalmente das redes cujo desempenho é avaliado pelo RMS é o *Levenberg-Marquardt* (LM), desenvolvido para minimizar funções que são somas de quadrados não lineares. O LM aproxima a matriz Hessiana a partir da Jacobiana no seu processo de otimização dos pesos, evitando um grande ônus computacional (HAGAN et al., 1996).

Figura 14 – Arquitetura de rede neural artificial MLP.



Fonte – Hagan et al. (1996)

O modelo de rede neural utilizado neste trabalho foi proposto por PELLETIER; MASSON; TAHAN(2016). Segundo Pelletier, Masson e Tahan, diversas configurações de rede foram testadas, sendo a que apresentou o melhor custo benefício entre complexidade e acurácia a seguinte rede: 4 neurônios e 2 camadas com funções de ativação do tipo tangente hiperbólica sigmoidal (primeira camada) e linear (segunda camada), treinadas pelo método de *backpropagation* com LM. Como entrada do modelo, contudo, foram utilizados nesta dissertação os resultados do módulo de redução de dimensionalidade, com todas as variáveis de entrada empregadas simultaneamente para o treino da rede.

4.2.4 Diagnóstico de Falhas

Para a etapa de diagnóstico de falhas, também foi empregado o conceito de informação mútua apresentado na Seção 4.2.2.2. Sua escolha deve-se ao fato da IM responder a seguinte questão: dada uma variável X, o quanto é possível saber sobre a variável Y? Partindo deste princípio, é possível estabelecer um análogo direto para o diagnóstico de falhas: dada uma variável qualquer do sistema SCADA, o quanto é possível saber sobre o resíduo r? Este novo indicador atua, portanto, como um índice de anomalia, sugerindo as variáveis mais provavelmente relacionadas às variações no resíduo.

A aplicação da IM com fins de diagnóstico dá-se da seguinte maneira:

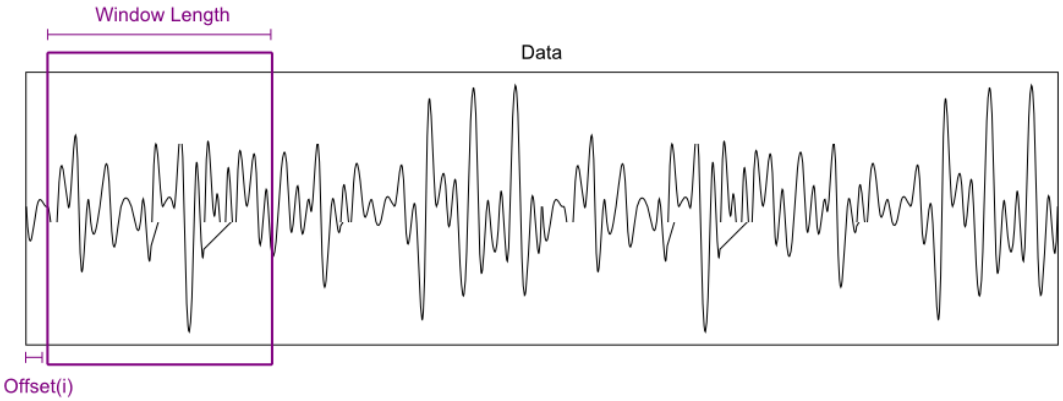
1. O resíduo é condicionado;
2. Caso uma falha tenha sido detectada, o dia em que se deu é salvo;
3. A IM entre o resíduo condicionado e cada uma das variáveis do sistema SCADA é calculada paralelamente para cada dia de dados disponíveis;
4. A partir do ordenamento da IM é possível determinar as variáveis que podem explicar a variação do resíduo no dia de ocorrência da falha e consequentemente desvios na produção.

O condicionamento do resíduo é feito a partir da aplicação do filtro de resposta finita passa baixa de *running sum* (SMITH, 2011). Seu emprego foi inspirado nos resultados de Butler, Ringwood e O'Connor (2013) e Karlsson (2015), que obtiveram êxito na detecção de falhas com baixa dissipação energética em rolamentos e gearboxes através da aplicação do resíduo acumulado. Uma janela móvel de largura $w = 144$ foi empregada e foi selecionada de forma a considerar um dia completo de dados.

Os últimos passos do método consistem no cálculo da informação mútua entre o resíduo condicionado e as variáveis do SCADA, bem como seu ordenamento para o dia em que houve a falha. Como a informação mútua é uma medida que envolve a distribuição de variáveis, uma janela significativa com comprimento $w = 144 \times 30 = 4320$ pontos foi considerada, o que equivale a uma janela de um mês de dados com amostragem de 10 minutos. Para cada dia, um novo valor de IM é gerado, ou seja, a janela move-se com um *offset* de $dt = 144$ pontos, conforme ilustrado na Figura 15, o que permite um monitoramento diário. Intervalos de dt desta ordem também foram utilizados por Yan, Liu e Gao (2012), em análise de entropia em séries temporais.

É importante destacar que a informação mútua explora associações de duas vias (YAO, 2003), ou seja ocorrência de X está associada a ocorrência de Y e vice-versa. Esta característica

Figura 15 – Janela móvel aplicada com um *offset* de $dt = 144$ pontos e janela de comprimento $w = 4320$ pontos.



Fonte – <https://kevinsprojects.wordpress.com>. Acesso em 30/08/2016

é fundamental ao explorar um banco de dados em busca de informações, pois necessariamente procuram-se indícios de relações entre as variáveis. O caso de ocorrência de uma falha é ainda mais específico, pois envolve a identificação de uma relação momentânea (local) entre as variáveis do SCADA e o sinal do resíduo condicionado. Espera-se, portanto, a maximização da IM entre o resíduo condicionado e a variável SCADA candidata à raiz do problema para o momento da falha.

5 Estudos de Caso

Neste capítulo, serão apresentados resultados da aplicação da metodologia de detecção e diagnóstico proposta para a aplicação sobre dados reais de aerogeradores. Porém, antes da aplicação da metodologia, uma seção discorrendo sobre a origem e característica dos dados empregados na análise é apresentada. Em seguida, dois estudos de caso são analisados em detalhe. Os estudos de caso selecionados consistem em:

- **Caso 1:** Aerogerador com regulação cuja causa raiz foi uma falha no sistema de controle;
- **Caso 2:** Aerogerador com regulação cuja causa raiz foi uma falha no sensor de temperatura ambiente, utilizado como proteção contra operação em condições extremas.

Casos de regulação de potência indevida indicam que o aerogerador produziu menos energia do que deveria para uma dada velocidade do vento. Pretende-se, com estes estudos de caso, validar não só a metodologia de detecção automática de falhas, mas também a inédita metodologia de diagnóstico proposta. Para isso, ela será aplicada a falhas reais similares, porém com causas raiz distintas.

5.1 Descrição dos Dados

Os dados utilizados nos estudos de caso aqui apresentados são provenientes de um sistema SCADA com cerca de 300 variáveis à disposição do operador. Inicialmente, todas elas foram consideradas para análise. Contudo, ficou claro que muitas destas variáveis não são de interesse para o monitoramento de performance, uma vez que são incapazes de indicar ou afetar a condição de operação do aerogerador. Com respeito às variáveis velocidade do vento e potência foram considerados os seguintes estatísticos: médias, mínimos, máximos e variância. No entanto, com respeito às demais variáveis considerou-se única e exclusivamente a média aritmética. Com isso, restaram as 54 variáveis listadas na Tabela 6 do Apêndice A.

Estas variáveis, empregadas no restante das análises, dizem respeito aos mais diversos subsistemas eletromecânicos dos aerogeradores, bem como suas condições ambientais de operação. Vale ressaltar que, por razões de confidencialidade, todos os dados relativos à produção serão normalizados pela potência nominal do aerogerador, resultando em comparativos de produção e erros em escala adimensional entre zero e um.

Os dados utilizados neste trabalho são provenientes de uma central eólica comercial que não pode ser identificada por questões de confidencialidade.

5.2 Resultados e Discussões

5.2.1 Garantia de Qualidade

A metodologia proposta por Moraes, com as adaptações aqui sugeridas, foi aplicada exclusivamente às séries temporais de velocidade média do vento e potência ativa. As demais variáveis do sistema SCADA não foram submetidas ao processo de garantia de qualidade. Sua aplicação, contudo, merece ressalvas: nem o teste de verificação temporal de curto prazo, nem o teste de consistência espacial foram aplicados. O teste de verificação temporal de curto prazo não foi aplicado, pois o número de amostras rechaçadas já foi grande o suficiente para a verificação de curtíssimo prazo, indicando a necessidade de modificações no mesmo como mostrado adiante. Já o teste de consistência espacial não chegou a ser completamente implementado devido a necessidade de ter-se um modelo de dados eficiente para todos os aerogeradores do parque com uma classificação segundo sua proximidade, o que não chegou a ser concretizado. Assim, a quantidade de dados detectados como potencialmente anômalos pelos testes remanescentes é apresentada na Tabela 3.

Tabela 3 – Porcentagem de dados potencialmente anômalos considerando a velocidade, a potência ativa e ambos, de acordo com metodologia de garantia de qualidade dos dados.

Caso	Velocidade	Potência Ativa	Total
Caso 1	$2,50\%$	$2,54\%$	$3,22\%$
Caso 2	$7,26\%$	$6,48\%$	$7,67\%$

Fonte – Próprio autor.

O número de dados potencialmente anômalos detectados pelo método de garantia de qualidade aplicado aos dados de vento foi de $2,50\%$ para o caso 1, o que é considerado elevado. Ao avaliar o resultado para a potência ativa do mesmo caso, o número é ainda maior, sendo quase $7,3\%$ dos dados tidos como potencialmente anômalos. O caso 2 apresenta resultados ainda maiores se comparado ao primeiro. Contudo ele também apresenta um maior número de dados faltantes. Para o caso 2, o número de intervalos de 10 minutos sem algum dado de potência ou velocidade corresponde a $6,35\%$ de toda a série, contra apenas $1,52\%$ no primeiro caso, ou seja, um valor quatro vezes maior. Subtraindo esta diferença, as porcentagens de dados classificados como potencialmente anômalos tornam-se equiparáveis. Vale mencionar que o critério aqui empregado é mais estrito que o critério de Tukey, usualmente empregado na análise exploratória de dados. Este último tipicamente classifica como anômalos $0,7\%$ dos dados em

casos de distribuição simétricas.

A análise conjunta dos indicadores de garantia de qualidade para todo o conjunto de dados rechaça um total de $3,22\%$ e $7,67\%$ dos dados para os casos 1 e 2 respectivamente. Proporcionalmente, o caso 1 possui uma menor interseção no conjunto de dados classificados como anômalos para a velocidade e a potência ativa, diferentemente do segundo caso onde o total de dados rechaçados encontra-se próximo numericamente dos valores individuais de cada variável, indicando uma maior concordância entre as classificações.

O teste dos limites apresentou-se como sendo o teste que mais classificou dados como potencialmente anômalos em ambos os casos, especialmente o critério que se utiliza do desvio padrão. Como a potência ativa varia (entre v_{cut-in} e v_{rated}) geralmente com o quadrado da velocidade do vento, é compreensível que considerando tal critério mais dados de potência resultem como potencialmente anômalos. Por esta razão, os resultados aqui apresentados não consideraram o teste dos limites para o desvio padrão da potência ativa.

Nenhum dos dados rechaçados na etapa de garantia de qualidade foi utilizado na modelagem de curva de potência a fim de garantir um modelo de operação normal fidedigno. Contudo, eles não foram excluídos da análise, tendo sido empregados nas etapas subsequentes.

5.2.2 Redução de Dimensionalidade

Tanto a PCA quanto a informação mútua foram aplicadas visando extrair o máximo de informação das 54 variáveis do SCADA analisadas, com o mínimo de esforço computacional para a elaboração de modelos de curva de potência. A aplicação da PCA no conjunto de dados resultou em 14 componentes principais, o que equivale a uma redução de 74% no volume de dados. Foi utilizado como critério de corte a retenção de 95% da variância dos dados originais, para os dois estudos de caso.

Já a aplicação do critério de maximização da informação mútua para a redução de dimensionalidade resultou em um ranking das variáveis do SCADA. As variáveis selecionadas em cada caso para a elaboração de CPs estão na Tabela 4. Foram selecionadas, apesar da diferença na prioridade em cada caso, a velocidade média do vento no segundo anemômetro, a corrente nas fases L2 e L3 para os casos 1 e 2 respectivamente e a potência reativa.

Note-se que não foram selecionadas as variáveis com os três maiores valores de IM e uma das razões foi a remoção de variáveis com comportamento similares, como correntes em diferentes enrolamentos do gerador ou velocidade de vento em um segundo anemômetro. Tampouco, foi feita a determinação do número de variáveis a serem selecionadas. Essencialmente, para consistência da metodologia objetiva proposta até o momento, um critério de seleção das

Tabela 4 – Variáveis selecionadas a partir do ranking da Informação Mútua entre as variáveis do SCADA e Potência Ativa.

Caso 1		Caso 2	
Posição	Variável	Posição	Variável
1	Corrente fase L2	1	Velocidade (anem. 2)
4	Velocidade (anem 2)	2	Corrente fase L3
6	Potência reativa	6	Potência reativa

Fonte – Próprio autor.

variáveis a partir do ranking deveria ter sido adotado. Isto só não foi feito em virtude dos resultados preliminares da modelagem das curvas de potência utilizando as variáveis segundo a Tabela 4, apresentados adiante na Seção 5.2.3.1.

5.2.3 Detecção de Falhas

5.2.3.1 Modelagem de Curva de Potência

A modelagem de curvas de potência se deu tanto a partir de modelos univariáveis[1] como a partir de modelos mais complexos que se valem de múltiplas variáveis[2]. Para tanto, os dados foram segmentados em três conjuntos amostrais distintos.

O primeiro deles é o conjunto amostral de treinamento (ST), composto apenas por dados:

- Classificados como bons pelo processo de garantia de qualidade;
- Obtidos do aerogerador quando em operação;
- Sem registro de ocorrência de falhas no livro de metadados para o período em questão.

Estes critérios não excluem a possibilidade de que dados potencialmente anômalos sejam utilizados para o treino ou calibração do modelo de CP, porém a minimiza. Em seguida, para cada caso, são tomados dois conjuntos amostrais para a validação do modelo de CP selecionados de forma subjetiva: um contendo a falha (regulação de potência) tal qual descrito no livro de metadados e outro livre de tal falha, denominados S1 e S2 respectivamente.

Cabe destacar que os modelos foram inicialmente treinados sobre ST (conjunto livre de falhas) com o objetivo de que os modelos elaborados reproduzissem o funcionamento normal do

[1] Polinômios do nono grau e quinto grau (P_9 e P_5), redes neurais (NN) e curva do fabricante (P_{man})
[2] Linear segmentado (LS) e redes neurais (NN)

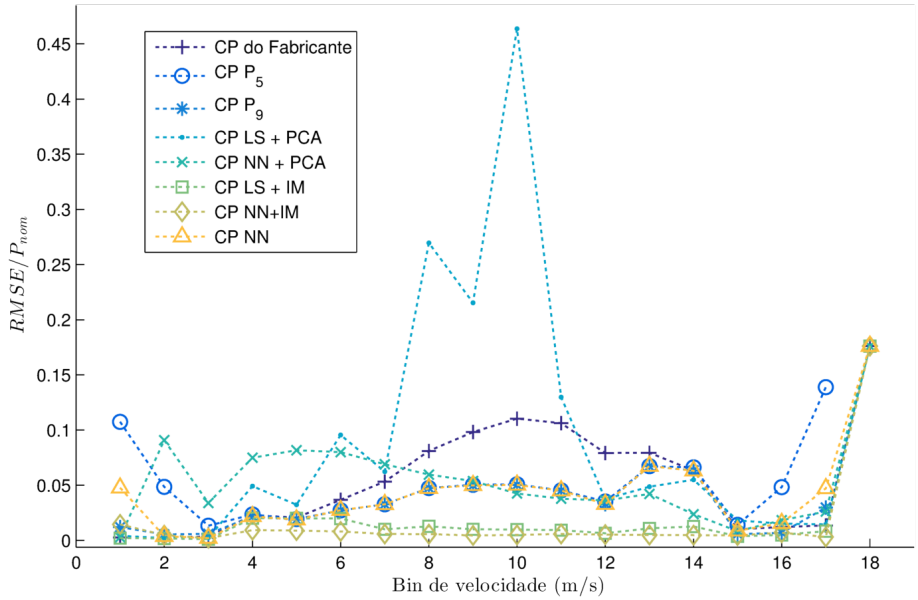
aerogerador. A validação do modelo ocorre sobre os conjuntos amostrais S1 e S2, de forma a verificar a habilidade na etapa de treinamento em ambas as circunstâncias.

Todos os conjuntos amostrais selecionados (ST, $S1$ e $S2$) possuem o equivalente a 10% do número total de dados disponíveis, resultando para o caso 1 em um conjunto com 5916 amostras e 3942 amostras para o caso 2. Procurou-se com isto garantir um número suficiente de dados para se obter uma curva de potência representativa de toda faixa de variação de velocidade e potência.

Como maneira de comparar todos os modelos propostos, foi adotado o erro RMS para a diferença entre a potência estimada pelos modelos de CP e a potência observada nas amostras S1 e S2. A seguir, são apresentados os resultados comparativos entre os modelos de CP para os casos 1 e 2, primeiramente para o conjunto livre de falhas (S2) e em seguida para o conjunto com a falha (S1).

A comparação do erro RMS por bin de velocidade de 1m/s sobre o conjunto de amostras S2 para o caso 1 é apresentada na Figura 16. É observado que, para velocidades entre a velocidade de arranque e a velocidade nominal, os modelos de CP elaborados com redes neurais e segmentação linear com variáveis selecionadas a partir da informação mútua apresentaram erros RMS quase que constantes a $0,01$ em toda a faixa de velocidade, enquanto todos os demais apresentaram um máximo global, com destaque ao modelo LS+PCA, cujo máximo de $0,45$ é muito superior ao dos demais modelos. Esta discrepância é ocasionada pela ineficiência do modelo em estimar valores de potência para períodos com intercorrências ou manutenção.

Figura 16 – Erro RMS por bin de velocidade de 1m/s para a diferença entre cada um dos modelos considerados e a potência observada na amostra S2 para o caso 1.



Fonte – Próprio autor.

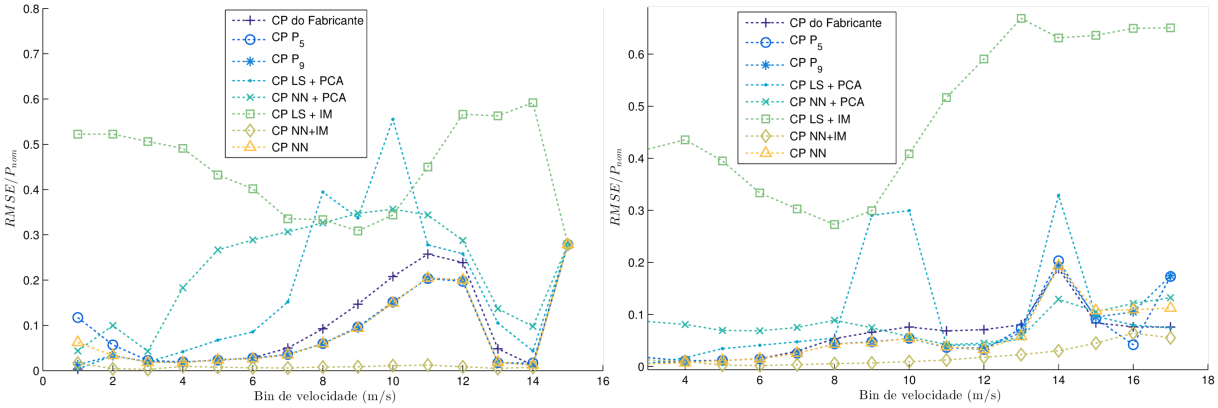
O modelo de CP polinomial de quinto grau, assim como o modelo polinomial de nono grau, apresenta erros RMS elevados na região de velocidade acima de 16m/s. Na realidade, os modelos polinomiais tendem a não estimar com acurácia valores de potência em regiões cujo número de amostras na etapa de treinamento não foi significativo, superestimando seus valores. Já para velocidades abaixo da velocidade de arranque, o modelo de CP P_9 não apresenta uma elevação do erro RMS, diferentemente do modelo de CP P_5.

Apesar da análise dos modelos em uma amostra livre de falhas ter sido exemplificada com dados do caso 1, o mesmo é válido para o caso 2. Isto se deve tanto à similaridade de condições operacionais dos aerogeradores (situados no mesmo parque eólico) quanto ao fato de serem do mesmo modelo.

Um panorama global do efeito da falha no erro RMS dos modelos analisados é dado nas Figuras 17 para os casos 1 e 2 respectivamente. Nota-se que o modelo LS com a informação mútua apresentou erros elevadíssimos da ordem de 50% da potência nominal da máquina. Ao se avaliar a CP disponibilizada no Apêndice B para os dois casos, é possível observar que os maiores desvios são provenientes dos pontos em que a máquina estava parada, modo de operação não utilizado durante o treinamento dos modelos.

Figura 17 – Erro RMS por bin de velocidade de 1m/s para a diferença entre cada um dos modelos considerados e a potência observada na amostra S1 (falha).

(a) Caso 1 (b) Caso 2



Fonte – Próprio autor.

Pela análise dos gráficos das Figuras 17a e 17b, o modelo que minimiza o erro RMS é o de redes neurais com a IM. Contudo, modelos com base na IM ajustam tão bem o comportamento da potência que terminam por simular o comportamento da falha. Isto ocorre porque eles são elaborados utilizando não só informações de variáveis ambientais como a velocidade do vento, mas também informações sobre a máquina, como potência reativa e corrente no enrolamento do gerador. Desta forma, tais modelos não podem ser selecionados com o objetivo de detectar variações na performance, pelo menos não utilizando o resíduo como métrica de falha uma vez

que a geração do resíduo no período com falha é mínima nestes casos. Como o emprego das variáveis com maior IM já resultou em resultados insatisfatórios, não foi elaborado nenhum critério objetivo para a seleção das variáveis ordenadas de acordo com o ranking da IM.

Por estas razões, resta apenas considerar os seguintes modelos de CP: polinomiais, de redes neurais e do fabricante, todos com apenas a velocidade do vento como variável de entrada. Os modelos polinomiais apresentaram um erro RMS inferior ao do fabricante na região da CP com maior probabilidade de falhas (região 1 da Figura 12), além disso, os modelos apresentam um RMS crescente justamente na faixa de velocidades em que elas se dão. Isto os tornam modelos propensos a baixas taxas de falsos alarmes e alta capacidade de detecção de variações na performance. Dentre os modelos polinomiais, o de nono grau é o que apresenta menores distorções de potência para valores de velocidade nos extremos da escala, embora ainda sejam elevadas. Similar ao desempenho da CP polinomial de nono grau foi o desempenho do modelo NN. Contudo, a capacidade de generalização da rede mostrou-se superior para o caso 2 (onde um menor número de dados de vento para a região de alta velocidade estava disponível na etapa de treinamento do modelo) se comparado com o primeiro caso.

Desta forma, foram selecionados como modelos de CP referenciais o modelo polinomial de nono grau (P_9) para o caso 1, e o modelo de baseado em redes neurais com apenas o vento como variável de entrada (NN) para o caso 2.

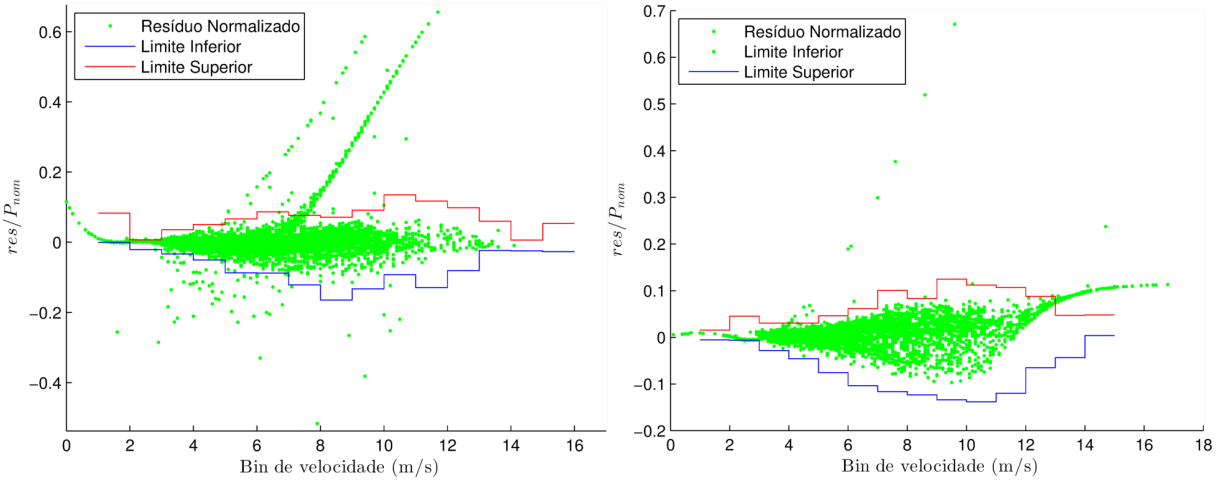
5.2.3.2 Estabelecimento de Limiares de Falha do Resíduo

Como resultado da etapa de modelagem de curva de potência, o modelo polinomial de nono grau foi utilizado como referência para o cálculo dos resíduos utilizados na detecção de falha do caso 1 e o modelo de redes neurais para o caso 2. O critério de detecção, por consistência da metodologia proposta, foi o mesmo utilizado para o cálculo das bandas limites do *box-plot*, sugerido por Hubert e Vandervieren (2008).

As bandas limites foram calculadas com base nos dados de treino e em seguida graficadas juntamente com os dados de teste da amostra S1 (contendo o período com falha) conforme Figura 18. É possível notar, para o caso 1 retratado na Figura 18a, dois filetes proeminentes acima da banda superior, indicando uma grande quantidade de pontos com falhas. O primeiro deles, localizado mais à esquerda, nada mais é do que o resíduo para velocidades acima da velocidade de arranque onde a turbina encontrava-se parada seja pelo acionamento de um alarme ou para ações de manutenção. Este filete pode ser evitado com a utilização de dados do SCADA referentes aos alarmes da máquina, que não foram explorados neste trabalho. O filete mais à direita se refere à regulação de potência e a medida que a velocidade do vento aumenta a magnitude do resíduo também o faz.

Figura 18 – Gráfico do resíduo normalizado das amostras S1 pela velocidade de vento (m/s), com as bandas de falha superior e inferior por bin de velocidade de 1m/s.

(a) Caso 1 (b) Caso 2



Fonte – Próprio autor.

Foi verificado para o caso 1 que 100% dos pontos onde houve de fato regulação de potência foram detectados. Para o caso 2 da Figura 18b, a regulação de potência não é tão proeminente como no caso 1 uma vez que ela se deu a um valor equivalente a 90% da potência nominal e a frequencia de incidência de vento com velocidades próximas ou acima da nominal é bastante inferior se comparada às demais. Desta forma, se tem uma combinação de falha com baixa taxa de incidência e que se manifesta comedidamente no resíduo, gerando uma diferença máxima de $0,1P_{nom}$. Além disso, a região de 9m/s a 13m/s corresponde à região onde ocorre a troca de modo de operação da máquina, uma vez que o aerogerador utiliza o *active-stall* como estratégia de regulação de potência, justificando uma maior abertura das bandas limites para esta região.

Visando comparar os benefícios dos modelos de curvas de potência propostos em detrimento da utilização da curva do fabricante para a geração de resíduo, foi elaborada a Tabela 5. Nela, considerando a classificação dos dados pelo processo de garantia de qualidade como referência, as classificações dos pontos pelo módulo de detecção de falhas utilizando-se da curva de potência polinomial de nono grau P_9 e curva de potência do fabricante P_{man} são comparadas para o estudo de caso 1 e curva de potência baseada em redes neurais com a velocidade média do vento como entrada e a curva de potência do fabricante para o caso 2.

Os resultados da Tabela 5 mostram que há uma alta incidência de falsos positivos e de forma menos proeminente de falsos negativos. Para o caso 1, o valor dos falsos positivos foi $5,25\%$ enquanto que no caso 2 foi de $3,02\%$. Esta é uma clara indicação que um maior esforço na etapa de garantia de qualidade é necessário, mesmo que os resultados já representem para cada caso uma redução de 300% e 55% com relação à CP do fabricante respectivamente. Tornar

Tabela 5 – Cruzamento de dados do sistema de detecção de falhas com o de garantia de qualidade, para o modelo de CP polinomial e do fabricante.

	Caso 1		Caso 2	
	P_9	P_{man}	NN	P_{man}
Verdadeiro Negativo	91,49	78,6	95,28	93,3
Falso Negativo	1,87	2,05	1,60	1,60
Falso Positivo	5,25	18,14	3,02	5,00
Verdadeiro Positivo	1,39	1,21	0,10	0,10

Fonte – Próprio autor.

a garantia de qualidade mais robusta beneficiará a metodologia automática de detecção de falhas na redução dos índices de classificação errôneas de falha.

Outro fator importante a ser destacado é que, enquanto para o caso 1 o modelo proposto está mais bem alinhado com relação à garantia de qualidade que a curva do fabricante, para o caso 2 o modelo proposto e a curva do fabricante equiparam-se. No caso 1, houve um aumento de 14% na classificação dos dados classificados como bons e um aumento de $1,21\%$ para $1,39\%$. No caso 2, o único ganho com relação à CP do fabricante foi a confirmação de $1,98\%$ dos falsos positivos como dados bons.

Desta forma, tem-se que bandas limites mais próximas devem ser estabelecidas para valores de velocidade de vento mais elevados de forma a aumentar a capacidade de detecção do modelo. Apesar de similar ao caso 1, a falha no caso 2 é menos intensa e se manifesta através de variações residuais pequenas.

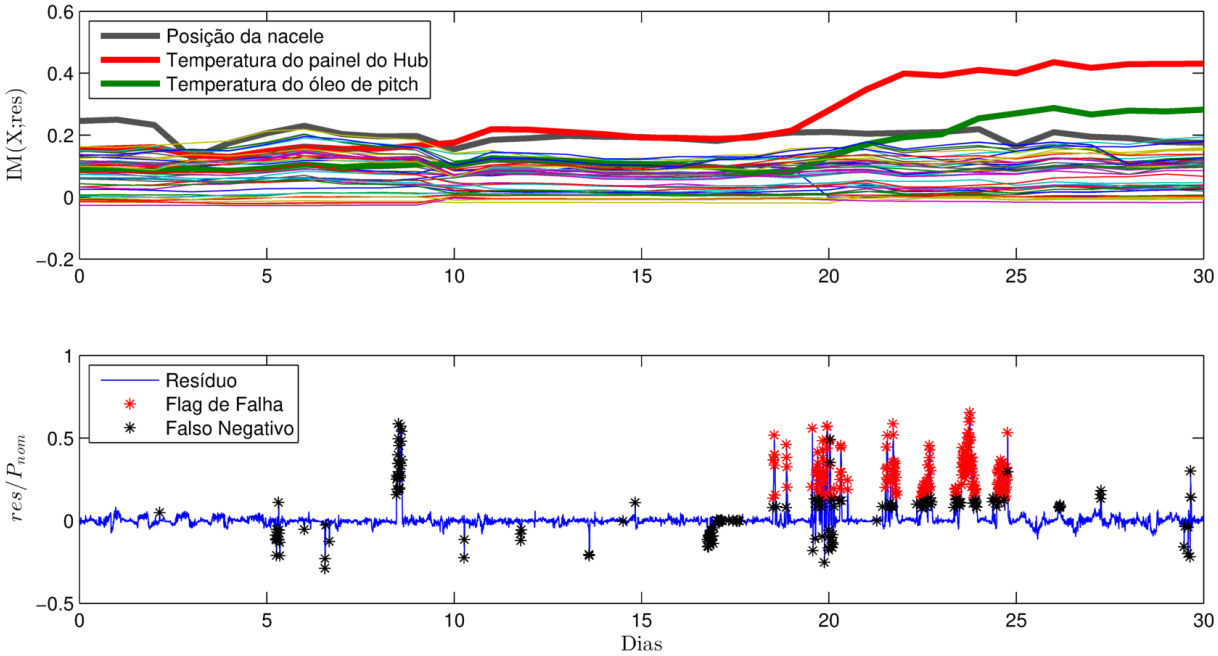
5.2.4 Diagnóstico de Falhas

A ideia por trás da metodologia de diagnóstico de falhas proposta é atuar como um índice de anomalia a ser consultado quando houver a detecção de uma falha. Desta forma, os cálculos da informação mútua entre o resíduo condicionado através da soma cumulativa e as variáveis do SCADA devem ser realizados de forma frequente.

Para o caso 1, o novo índice de anomalia é apresentado na parte superior da Figura 19 para um período de 30 dias. A informação mútua entre o resíduo condicionado por uma janela de soma cumulativa e cada uma das variáveis do sistema SCADA consideradas na análise foi calculada para cada dia. Suas legendas foram suprimidas de forma a tornar o gráfico legível. Apenas as variáveis que se destacaram tiveram as legendas inseridas e traços realçados. Ao confrontar o índice de anomalia com o período de resíduos elevados determinados pelo processo

de detecção de falha, tal qual ilustrado na parte inferior da Figura 19, algumas variáveis reagem diferentemente das demais. É o caso da temperatura do painel do hub e temperatura do óleo de pitch. Isto dá indícios de que estas variáveis estejam relacionadas ao modo de falha. No caso da regulação de potência, é comum a bomba do sistema de pitch trabalhar continuamente de forma a manter a pressão nos atuadores em níveis elevados para que possam ser manobrados, sobreaquecendo o óleo que por ela passa. Além disso, o computador do hub é localizado fisicamente próximo aos mesmos atuadores. De acordo com o livro de registros, é fato que a causa raiz foi uma falha no sistema de controle que afetou diretamente o sistema de pitch, responsável pela regulação de potência, como apontado pelo indicador proposto.

Figura 19 – Caso 1 - Gráfico superior: informação mútua entre o resíduo condicionado e variáveis do SCADA com destaque para as variáveis mais prováveis de serem a causa raiz da falha. Gráfico inferior: resíduo entre o modelo P_9 de CP e os dados observacionais de potência ativa com regulações detectadas e não detectadas.



Fonte – Próprio autor.

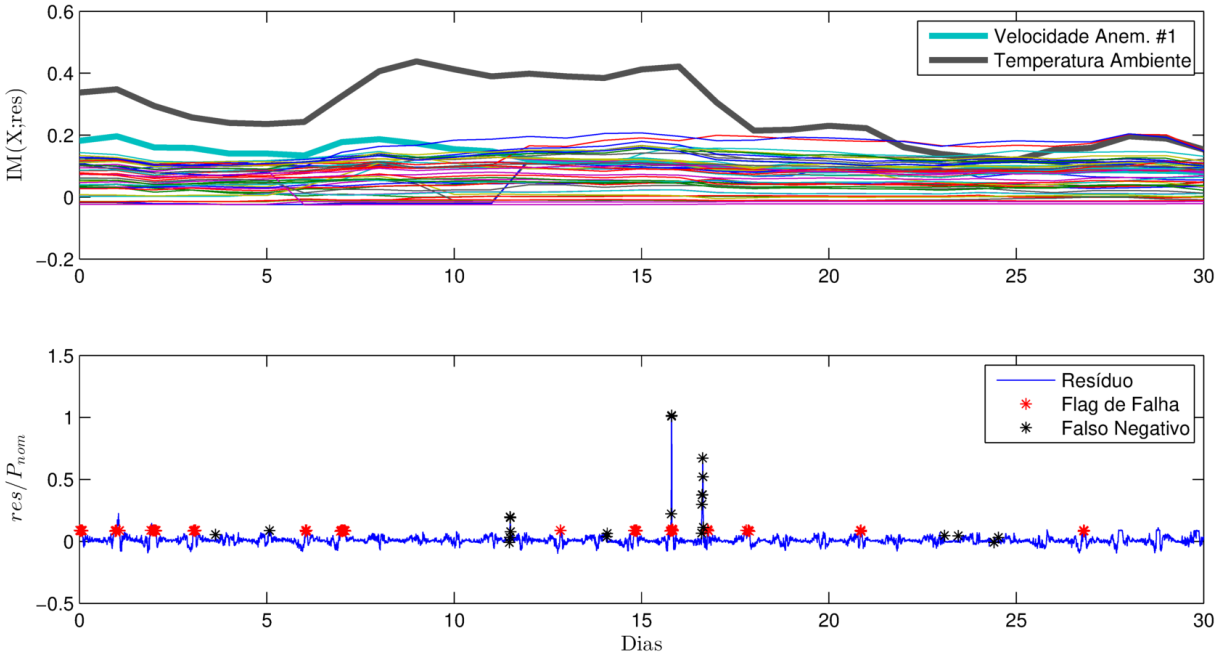
Um breve comentário deve ser feito sobre a posição da nacele. Esta variável aparece como primeiro colocado no índice de anomalia no gráfico superior da Figura 19 para os primeiros cinco dias de análise. A posição da nacele é uma variável circular e, quando próxima dos limites 0° ou 360°, ocorrem variações escalares bruscas ainda que a variação vetorial seja pequena. Estas variações, verdadeiros "pulos", apesar de não representativas de falhas na performance, ocorreram conjuntamente com a flutuação do resíduo fazendo com que esta variável atingisse o topo do ranking. Estas flutuações entre 0° e 360° acarretam em alterações na pdf da posição da nacele, que ocorrem concomitantemente com variações no próprio resíduo, ocasionadas por falhas anemométricas devidamente detectadas pelo processo de garantia de qualidade. Por conta

disto a informação mútua se eleva.

O estudo de caso número 2 é similar ao primeiro. Contudo, a causa raiz da regulação está relacionada ao mecanismo de defesa do aerogerador de não operar com carga máxima em condições ambientes extremas, como temperatura ambiente superior a 40 °C.

A Figura 20 também mostra um intervalo de 30 dias tanto para o gráfico do índice de anomalia quanto para o do resíduo. Verifica-se que duas variáveis possuem informação mútua (com relação ao resíduo condicionado) superior às demais. Elas são a velocidade do vento lida pelo anemômetro #1 e a temperatura ambiente. Isto significa que alterações na distribuição destas variáveis podem ter ocasionado as variações no resíduo. No entanto, a modificação nas leituras do anemômetro como causa raiz da regulação foram descartadas, uma vez que para o período com falhas ele apresentou um congelamento de leituras, já detectado no processo de garantia de qualidade. Restam, portanto, as variações da temperatura ambiente como alternativa a justificar as variações no resíduo. Ocorre que o sensor de temperatura ambiente estava de fato danificado, registrando valores de temperaturas até 400% superiores aos reais e que a máquina, como mecanismo de proteção, limitou sua produção de forma a preservar os componentes principais. O mecanismo de diagnóstico proposto detectou não só a variação na produção como também a variável associada a real causa raiz do problema.

Figura 20 – Caso 2 - Gráfico superior: informação mútua entre o resíduo condicionado e variáveis do SCADA com destaque para as variáveis mais prováveis de serem a causa raiz da falha. Gráfico inferior: resíduo entre o modelo *NN* de CP e os dados observacionais de potência ativa com regulações detectadas e não detectadas indicadas.



Fonte – Próprio autor.

6 Conclusões e Perspectivas

6.1 Conclusões

O sistema de detecção e diagnóstico de falhas na performance de aerogeradores com base em dados do sistema SCADA proposto foi implementado e avaliado utilizando-se dados reais com intervalo de integração de 10 minutos. O sistema é composto por quatro módulos: garantia de qualidade, redução de dimensionalidade, detecção e diagnóstico de falhas.

Mostrou-se a bem sucedida adaptação do trabalho de Moraes (2014) com foco na garantia de qualidade de dados de vento em superfície para também dados anemométricos e dados de potência de aerogeradores. Mostrou-se, também, que a substituição do método não paramétrico (aplicado na etapa objetiva do processo) conferiu robustez à metodologia. No entanto, ela aumentou a quantidade de dados rechaçados. Tais rechaços se dão especialmente nos limites inferiores dos intervalos de potência e velocidade, ocasionando um alto índice de falsos positivos.

No campo da redução de dimensionalidade, a utilização da análise de componentes principais e informação mútua resultou em modelos com comportamentos desde extremamente satisfatórios com erro RMSE de $0,01\,P_{nom}$ até modelos com erros na ordem de $0,5\,P_{nom}$, como nos modelos NN+IM (S1) e LS+IM (S1) para os casos 1 e 2 respectivamente. Além disso, houve uma redução de 75% do volume de dados utilizados quando aplicada a PCA, chegando a uma redução de $94,4\%$ com a aplicação da IM. Contudo, modelos com dados de entrada de dimensionalidade reduzida mostraram-se inapropriados para o monitoramento de performance, uma vez que reproduziram as falhas apresentadas pelos aerogeradores. Desta forma, o monitoramento foi levado adiante através da utilização da curva de potência polinomial e de redes neurais com apenas o vento como dado de entrada.

O sistema de detecção de falhas (através do método não paramétrico de *boxplot*) resultou na detecção de até 100% dos dados com regulação para o caso 1 e de $47,18\%$ para o caso 2, em que a magnitude do resíduo não ultrapassou o limite de $0,1\,P_{nom}$. O índice de falsos positivos ainda foi elevado devido principalmente aos ajustes necessários no mecanismo de garantia de qualidade de dados, especialmente daqueles ajustes referentes à potência ativa do aerogerador.

Por fim, o novo método elaborado com base na informação mútua utilizado para realizar o diagnóstico de falhas provou-se eficaz na localização de sua causa raiz. Para o mesmo problema de regulação, o método foi capaz de detectar quais variáveis a ocasionaram: aquelas relacionadas

ao sistema de pitch para o caso 1 e à elevação da temperatura ambiente para o caso 2.

6.2 Recomendações

Os resultados mostram que a adaptação do trabalho de Moraes (2014) ainda pode ser aperfeiçoada, de forma a rechaçar uma porcentagem menor de dados, especialmente nos limites inferiores dos intervalos de potência e velocidade. Além disso, é essencial expandir o trabalho ao ponto em que todos os dados do sistema SCADA passem pelo processo de garantia de qualidade.

No âmbito da redução de dimensionalidade, um novo estudo pode ser realizado avaliando a eficácia da seleção apenas de variáveis ambientais como entrada dos modelos multidimensionais de curvas de potência (PELLETIER; MASSON; TAHAN, 2016). Além disso, mecanismos que garantam uma quantidade mínima de pontos por bin de velocidade podem ser implementados de forma que os modelos sejam treinados por toda faixa de velocidade.

No que se refere à elaboração das curvas de potência, é necessário enfatizar que os dados observacionais de potência não foram corrigidos por sua densidade conforme sugerido pela norma IEC 61400-12-1. Assim, deve-se considerá-la em trabalhos futuros sempre que possível, de forma a tornar as comparações com a CP do fabricante (tomada a uma densidade constante) mais coerentes.

No tocante à detecção de falhas, um aspecto desejável é o estreitamento das bandas limites para bins de velocidade mais elevada. Com isto, haveria um aumento da sensibilidade do método à detecção de falhas que afetem sutilmente a produção nestes intervalos de velocidade.

Por último, com respeito ao diagnóstico, uma grande adição ao trabalho seria quantificar a probabilidade de que uma determinada variável seja a causa raiz da falha. Para isto, um dos possíveis caminhos sugeridos se dá através da adição da relevância simétrica (MEYER; BONTEMPI, 2006), utilizada na classificação de anomalias genéticas.

Referências Bibliográficas

ABDI, H. A neural network primer. *Journal of Biological Systems*, v. 02, n. 03, p. 247–281, 1994. Disponível em: <http://www.worldscientific.com/doi/abs/10.1142/S0218339094000179>. (Citado na página 39).

BENNASAR, M.; HICKS, Y.; SETCHI, R. Feature selection using joint mutual information maximisation. *Expert Systems with Applications*, v. 42, n. 22, p. 8520 – 8532, 2015. ISSN 0957-4174. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0957417415004674>. (Citado na página 35).

BREEZE. *Evaluation of Power Performance Upgrades to Wind Turbines*. [S.l.], 2015. Disponível em: <www.breeze.com>. Acesso em: 26/06/2016. (Citado na página 38).

BURTON, T. et al. *Wind energy handbook*. [S.l.]: John Wiley & Sons, 2001. ISBN 0-471-48997-2. (Citado na página 18).

BUTLER, S.; RINGWOOD, J. V.; O'CONNOR, F. Exploiting scada system data for wind turbine performance monitoring. In: IEEE. *Control and fault-tolerant systems (systol), 2013 conference on*. 2013. p. 389–394. Disponível em: <http://eprints.maynoothuniversity.ie/6794/1/JRSCADA%20system.pdf>. (Citado nas páginas 16, 25, 27, 41).

COVER, T. M.; THOMAS, J. A. *Elements of Information Theory*. 2. ed. New York, NY, USA: Wiley-Interscience, 1991. ISBN 0-471-06259-6. Disponível em: <http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471241954.html>. (Citado nas páginas 35, 36).

DING, S. X. *Data-driven Design of Fault Diagnosis and Fault-tolerant Control Systems*. 1. ed. Springer-Verlag London, 2014. Disponível em: <http://www.springer.com/us/book/9781447164098>. (Citado na página 33).

FODOR, I. *A Survey of Dimension Reduction Techniques*. 2002. Disponível em: <https://e-reports-ext.llnl.gov/pdf/240921.pdf>. (Citado nas páginas 33, 34).

GRAVES, A. Wind operation & maintenance strategy - long-term ownership: Challenges & what to expect. In: . [S.l.]: Apresentado na AWEA, Texas, USA, 2014. Acesso em: 20/02/2016. (Citado nas páginas 25, 26).

GRAY, C. S. et al. A practical approach to the use of scada data for optimized wind turbine condition based maintenance. In: *EWEA Offshore Wind Amsterdam*. [s.n.], 2011. Disponível em: <http://www.uptime-engineering.com/downloads/Practical_Approach_SCADA_Wind_Maintenance_GRAY_etal.pdf>. (Citado na página 31).

GUO, P.; INFIELD, D.; YANG, X. Wind turbine generator condition-monitoring using temperature trend analysis. *Sustainable Energy, IEEE Transactions on*, v. 3, n. 1, p. 124–133, Jan 2012. ISSN 1949-3029. (Citado na página 25).

GWEC. *Global Wind Report - Annual Market Update*. [S.l.], 2015. Disponível em: <http://www.gwec.net/wp-content/uploads/vip/GWEC-Global-Wind-2015-Report_April-2016_22_04.pdf>. Acesso em: 25/04/2016. (Citado na página 15).

HAGAN, M. T. et al. *Neural network design*. PWS publishing company Boston, 1996. v. 20. Disponível em: <http://hagan.okstate.edu/NNDesign.pdf>. (Citado na página 40).

HANSEN, M. O. *Aerodynamics of wind turbines*. [S.l.]: Routledge, 2015. (Citado na página 18).

HAU, E. *Wind Turbines-Fundamentals, Technologies, Application, Economics.* [S.l.]: Springer, 2005. (Citado na página 19).

HUBERT, M.; VANDERVIEREN, E. An adjusted boxplot for skewed distributions. *Computational Statistics & Data Analysis*, v. 52, n. 12, p. 5186 – 5201, 2008. ISSN 0167-9473. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0167947307004434>. (Citado nas páginas 32, 49).

HUGHES, G. The performance of wind farms in the united kingdom and denmark. *Renewable Energy Foundation*, p. 48, 2012. Disponível em: <http://www.pfbach.dk/firma_pfb/gordon_hughes_wind_farm_perf_2012.pdf>. (Citado na página 15).

HYERS, R. et al. Condition monitoring and prognosis of utility scale wind turbines. *Energy Materials*, Maney Publishing, v. 1, n. 3, p. 187–203, 2006. Disponível em: <https://www.researchgate.net/profile/Jon_Mcgowan/publication/228873395_Condition_monitoring_and_prognosis_of_utility_scale_wind_turbines._Energy_Mater/links/0c96052e8fb7e082eb000000.pdf>. (Citado na página 25).

IEC. *Power performance measurements of electricity producing wind turbines*. 2005. (Citado nas páginas 21, 22).

IRENA. *Renewable Power Generation Costs in 2014*. [S.l.], 2014. Disponível em: <http://www.irena.org/DocumentDownloads/Publications/IRENA_RE_Power_Costs_2014_report.pdf>. Acesso em: 03/01/2016. (Citado na página 22).

ISERMANN, R. *Fault-Diagnosis Applications*. Springer Science + Business Media, 2011. Disponível em: <http://dx.doi.org/10.1007/978-3-642-12767-0>. (Citado na página 20).

JIMÉNEZ, P. A. et al. Quality assurance of surface wind observations from automated weather stations. *Journal of Atmospheric and Oceanic Technology*, American Meteorological Society, v. 27, n. 7, p. 1101–1122, Mar 2010. ISSN 0739-0572. Disponível em: <http://dx.doi.org/10.1175/2010JTECHA1404.1>. (Citado na página 31).

KARLSSON, D. *Wind Turbine Performance Monitoring using Artificial Neural Networks With a Multi-Dimensional Data Filtering Approach*. Dissertação (Mestrado) — Chalmers University of Technology, Sweden, 2015. 88. Disponível em: <http://publications.lib.chalmers.se/records/fulltext/221324/221324.pdf>. (Citado nas páginas 28, 41).

KUSIAK, A.; VERMA, A. Monitoring wind farms with performance curves. *Sustainable Energy, IEEE Transactions on*, IEEE, v. 4, n. 1, p. 192–199, Jan 2013. ISSN 1949-3029. (Citado nas páginas 16, 26).

LETZGUS, S. *SCADA-Data Analysis for Condition Monitoring of Wind Turbines*. Dissertação (Mestrado) — Chalmers University of Technology, Gothenburg, Sweden, 08 2015. Disponível em: <http://publications.lib.chalmers.se/records/fulltext/220577/220577.pdf>. (Citado nas páginas 25, 26, 28, 41).

LYDIA, M. et al. A comprehensive review on wind turbine power curve modeling techniques. *Renewable and Sustainable Energy Reviews*, v. 30, p. 452 – 460, 2014. ISSN 1364-0321. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1364032113007296>. (Citado nas páginas 25, 27).

MAGALHÃES, M. N. *Noções de probabilidade e estatística*. [S.l.]: EDUSP, 2005. ISBN 8531406773. (Citado na página 23).

MAIMON, O.; ROKACH, L. (Ed.). *Data Mining and Knowledge Discovery Handbook*. Springer Science + Business Media, 2010. Disponível em: <http://dx.doi.org/10.1007/978-0-387-09823-4>. (Citado na página 33).

MARVUGLIA, A.; MESSINEO, A. Monitoring of wind farms power curves using machine learning techniques. *Applied Energy*, v. 98, p. 574 – 583, 2012. ISSN 0306-2619. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0306261912003236>. (Citado na página 28).

MEYER, P. E.; BONTEMPI, G. On the use of variable complementarity for feature selection in cancer classification. In: SPRINGER. *Workshops on Applications of Evolutionary Computation*. 2006. p. 91–102. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.84.5883&rep=rep1&type=pdf>. (Citado na página 55).

MOBLEY, R. *An Introduction to Predictive Maintenance*. Elsevier Science, 2002. (Plant Engineering). ISBN 9780080478692. Disponível em: <https://books.google.com.br/books?id=SjqXzxpAzSQC>. (Citado na página 22).

MORAES, C. *Procedimento Objetivo para a Garantia da Qualidade de Dados Observacionais de Vento em Superfície no Litoral do Rio Grande do Norte*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, PPGEM, Av. Professor Moraes Rego, 1235 - Cidade Universitária, Fevereiro 2014. P. 87. (Citado nas páginas 16, 31, 32, 54, 55).

NILSSON, J.; BERTLING, L. Maintenance management of wind power systems using condition monitoring systems - life cycle cost analysis for two case studies. *IEEE Transactions on Energy Conversion*, v. 22, n. 1, p. 223–229, March 2007. ISSN 0885-8969. (Citado na página 15).

PAIVA, L.; RODRIGUES, V. C.; PALMA, J. Determining wind turbine power curves based on operating conditions. *Wind Energy*, v. 17, n. 10, p. 1563–1575, 2014. ISSN 1099-1824. Disponível em: <http://dx.doi.org/10.1002/we.1651>. (Citado nas páginas 31, 32).

PELLETIER, F.; MASSON, C.; TAHAN, A. Wind turbine power curve modelling using artificial neural network. *Renewable Energy*, v. 89, p. 207 – 214, 2016. ISSN 0960-1481. Disponível em: <http://www.sciencedirect.com/science/article/pii/S096014811530481X>. (Citado nas páginas 38, 40, 55).

PENG, H.; LONG, F.; DING, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 27, n. 8, p. 1226–1238, 2005. (Citado na página 35).

RAGHEB, M.; RAGHEB, A. M. *Wind turbines theory-the betz equation and optimal rotor tip speed ratio*. INTECH Open Access Publisher, 2011. Disponível em: <http://mragheb.com/Wind_turbines_theory_the_betz_equation_and_optimal_rotor_tip_speed_ratio.pdf>. Acesso em: 15/05/2016. (Citado na página 18).

REN21. *Renewables 2015: Global Status Report*. [S.l.], 2015. Disponível em: <http://www.ren21.net/status-of-renewables/global-status-report/>. Acesso em: 02/01/2016. (Citado na página 15).

SCHLECHTINGEN, M.; SANTOS, I. F. Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. *Mechanical Systems and Signal Processing*, v. 25, n. 5, p. 1849 – 1875, 2011. ISSN 0888-3270. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0888327010004310>. (Citado na página 25).

SERGENT, M. et al. Correct and incorrect use of multilinear regression. *Chemometrics and Intelligent Laboratory Systems*, v. 27, n. 2, p. 153 – 162, 1995. ISSN 0169-7439. Disponível em: <http://www.sciencedirect.com/science/article/pii/016974399580020A>. (Citado na página 37).

SHLENS, J. A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*, 2014. Disponível em: <http://arxiv.org/pdf/1404.1100.pdf>. Acesso em: 19/09/2015. (Citado nas páginas 33, 34).

SINGH, S.; BHATTI, T. S.; KOTHARI, D. P. Wind power estimation using artificial neural network. *J. Energy Eng.*, American Society of Civil Engineers (ASCE), v. 133, n. 1, p. 46–52, Mar 2007. ISSN 1943-7897. Disponível em: <http://dx.doi.org/10.1061/(ASCE)0733-9402(2007)133:1(46)>. (Citado na página 28).

SMITH, J. O. *Spectral audio signal processing*. [S.l.]: W3K, 2011. (Citado na página 41).

STOUFFER, K. et al. Guide to supervisory control and data acquisition (scada) and industrial control systems security. In: *SPIN*. [s.n.], 2006. Disponível em: <http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-82r1.pdf>. (Citado na página 21).

TUKEY, J. W. *Exploratory Data Analysis*. [S.l.]: Addison-Wesley, 1977. (Citado nas páginas 24, 32).

ULUYOL, O.; PARTHASARATHY, G. Multi-turbine associative model for wind turbine performance monitoring. In: *Proceedings of the Annual Conference of the Prognostics and Health Management Society, Minneapolis, MN, USA*. [s.n.], 2012. v. 2327. Disponível em: <http://www.phmsociety.org/sites/phmsociety.org/files/phm_submission/2012/phmc_12_029.pdf>. (Citado na página 16).

ULUYOL, O. et al. Power curve analytic for wind turbine performance monitoring and prognostics. In: *Annual conference of the prognostics and health management society*. [s.n.], 2011. Vol 2. Disponível em: <https://www.phmsociety.org/sites/phmsociety.org/files/phm_submission/2011/phmc_11_049.pdf>. (Citado nas páginas 16, 17, 26, 27, 37).

VEJEN, F. et al. *Quality Control of Meteorological Observations*. [S.l.], 2002. Disponível em: <http://www.smhi.se/hfa_coord/nordklim/task1/quality_control.pdf>. (Citado na página 31).

VERBOVEN, S.; HUBERT, M. LIBRA: a MATLAB Library for Robust Analysis. *Chemometrics and Intelligent Laboratory Systems*, v. 75, p. 127–136, 2004. Disponível em: <http://wis.kuleuven.be/stat/robust/Papers/LIBRA.pdf>. (Citado na página 32).

VIHAROS, Z. J. et al. "big data"initiative as an it solution for improved operation and maintenance of wind turbines. In: *European Wind Energy Association (EWEA) Conference*. [s.n.], 2013. p. 184–188. Disponível em: <http://eprints.sztaki.hu/7753/1/Viharos_184_2476933_ny.pdf>. (Citado na página 21).

WANG, K.; SHARMA, V. S.; ZHANG, Z. Scada data based condition monitoring of wind turbines. *Advances in Manufacturing*, Shanghai University, v. 2, n. 1, p. 61–69, 2014. ISSN 2095-3127. Disponível em: <http://dx.doi.org/10.1007/s40436-014-0067-0>. (Citado na página 28).

WANG, Y.; INFIELD, D. Supervisory control and data acquisition data-based non-linear state estimation technique for wind turbine gearbox condition monitoring. *Renewable Power Generation, IET*, v. 7, n. 4, p. 350–358, July 2013. ISSN 1752-1416. (Citado nas páginas 16, 25).

WILKINSON, M. et al. Comparison of methods for wind turbine condition monitoring with scada data. *Renewable Power Generation, IET*, v. 8, n. 4, p. 390–397, May 2014. ISSN 1752-1416. (Citado na página 25).

YAMPIKULSAKUL, N. et al. Condition monitoring of wind power system with nonparametric regression analysis. *Energy Conversion, IEEE Transactions*, v. 29, n. 2, p. 288–299, June 2014. ISSN 0885-8969. (Citado na página 28).

YAN, R.; LIU, Y.; GAO, R. X. Permutation entropy: a nonlinear statistical measure for status characterization of rotary machines. *Mechanical Systems and Signal Processing*, Elsevier, v. 29, p. 474–484, 2012. Disponível em: <https://www.researchgate.net/profile/Ruqiang_Yan/publication/257271852_Permutation_entropy_A_nonlinear_statistical_measure_for_status_characterization_of_rotary_machines/links/553abc350cf245bdd76448c5.pdf>. (Citado na página 41).

YANG, H.; MOODY, J. Feature selection based on joint mutual information. In: *Proceedings of international ICSC symposium on advances in intelligent data analysis*. [s.n.], 1999. p. 22–25. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.41.4424&rep=rep1&type=pdf>. (Citado na página 35).

YANG, W.; COURT, R.; JIANG, J. Wind turbine condition monitoring by the approach of scada data analysis. *Renewable Energy*, v. 53, p. 365 – 376, 2013. ISSN 0960-1481. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0960148112007653>. (Citado na página 25).

YANG, W.; JIANG, J. Wind turbine condition monitoring and reliability analysis by scada information. In: *Second International Conference on Mechanic Automation and Control Engineering (MACE)*. [S.l.: s.n.], 2011. p. 1872–1875. (Citado na página 21).

YAO, Y. Information-theoretic measures for knowledge discovery and data mining. In: *Entropy Measures, Maximum Entropy Principle and Emerging Applications*. Springer, 2003. p. 115–136. Disponível em: <http://www2.cs.uregina.ca/~yyao/PAPERS/information_measures.pdf>. (Citado na página 41).

ZAHER, A. et al. Online wind turbine fault detection through automated scada data analysis. *Wind Energy*, v. 12, n. 6, p. 574–593, 2009. ISSN 1095-4244. (Citado na página 28).

ZHANG, Z.-Y.; WANG, K.-S. Wind turbine fault detection based on scada data analysis using ann. *Advances in Manufacturing*, Shanghai University, v. 2, n. 1, p. 70–78, 2014. ISSN 2095-3127. Disponível em: <http://dx.doi.org/10.1007/s40436-014-0061-6>. (Citado na página 25).

Apêndice A - Variáveis do Sistema SCADA

Tabela 6 – Lista de variáveis do sistema SCADA filtradas.

#	Variável	Unidade
1	Avr.active power.	kW
2	Avr.curr.in phase L1.	A
3	Avr.curr.in phase L2.	A
4	Avr.curr.in phase L3.	A
5	Avr.reactive power.	kVAr
6	Avr.voltage phase L1.	V
7	Avr.voltage phase L2.	V
8	Avr.voltage phase L3.	V
9	Avg. relative wind direction 1	°
10	Avg. relative wind direction 2	°
11	Avr.wind speed 1.	m/s
12	Avr.wind speed 2.	m/s
13	Var.wind speed 1	m^2/s^2
14	Var.wind speed 2	m^2/s^2
15	Avr.blade A pressure.	bar
16	Avr.blade B pressure.	bar
17	Avr.blade C pressure.	bar
18	Avr.gear oil pressure.	bar
19	Avr.gen.speed.	rpm
20	Avr.hub inlet pressure.	bar
21	Avr.nacelle pos.	°
22	Avr.pos. blade A	°
23	Avr.pos. blade B	°
24	Avr.pos. blade C	°
25	Avr.rotor speed.	rpm
26	Avr.shaft br. pressure.	bar
27	Avr.yaw br. pressure.	bar
28	AvgTac84DownwindVibration	-
29	AvgTac84Vibration	-

Tabela 6 Continuação: Variáveis do Sistema SCADA

30	AvgTac85Rpm1	rpm
31	AvgTac85Rpm2	rpm
32	Avr.ambient temp.	°C
33	Avr.control panel temp.	°C
34	Avr.gear oil temp.	°C
35	Avr.hub panel temp.	°C
36	Avr.main panel temp.	°C
37	Avr.nacelle temp.	°C
38	Avr.ph.comp. panel temp.	°C
39	Avr.W1 temp.large gen.	°C
40	Avr.W2 temp.large gen.	°C
41	Avr.W3 temp.large gen.	°C
42	Avr.t.front gear bearing.	°C
43	Avr.t.front gen. bearing.	°C
44	Avr.t.rear gear bearing.	°C
45	Avr.t.rear gen. bearing.	°C
46	Avr.t.gear oil after exch.	°C
47	Avr.temp.main bearing 1	°C
48	Avr.temp.pitch oil.	°C
49	Avr.temp.tower base.	°C
50	Avr.temp.water a.cooler.	°C
51	Avr.temp.water b.cooler.	°C
52	Avr.thyristor temp.	°C
53	Avr.topbox temp.	°C
54	Avr.wind.temp.large gen.	°C

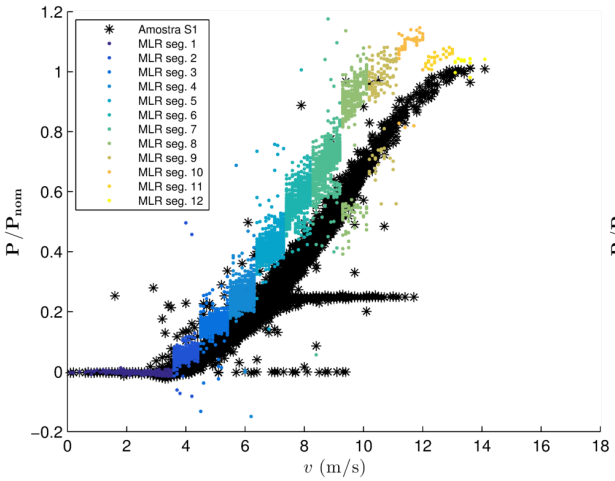
Fonte – Próprio autor.

Apêndice B - Curvas de Potência

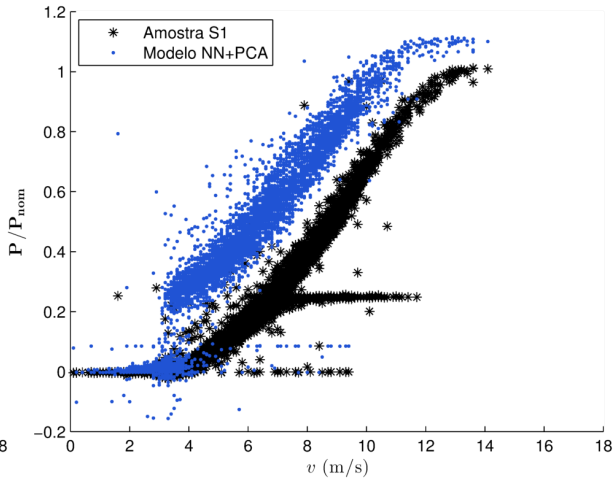
B.1 Caso 1

Figura 21 – Distorções da CP modelada utilizando dados provenientes da PCA para o caso 1.

(a) Modelo LS+PCA



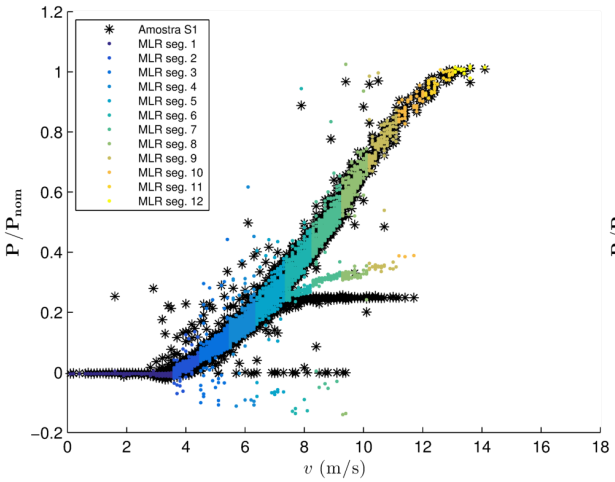
(b) Modelo NN+PCA



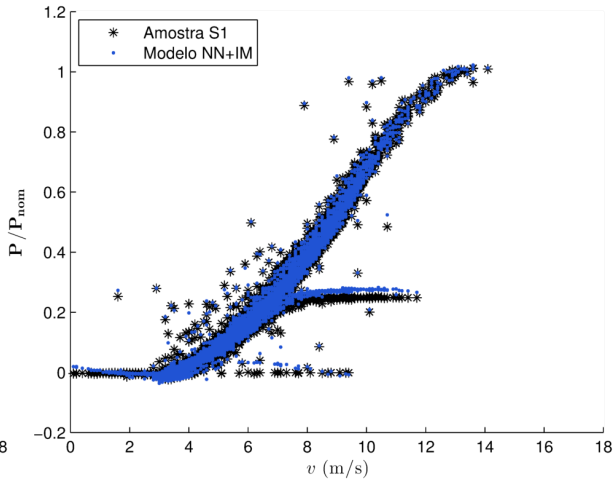
Fonte – Próprio autor.

Figura 22 – Distorções da CP modelada utilizando dados provenientes da IM para o caso 1.

(a) Modelo LS+IM



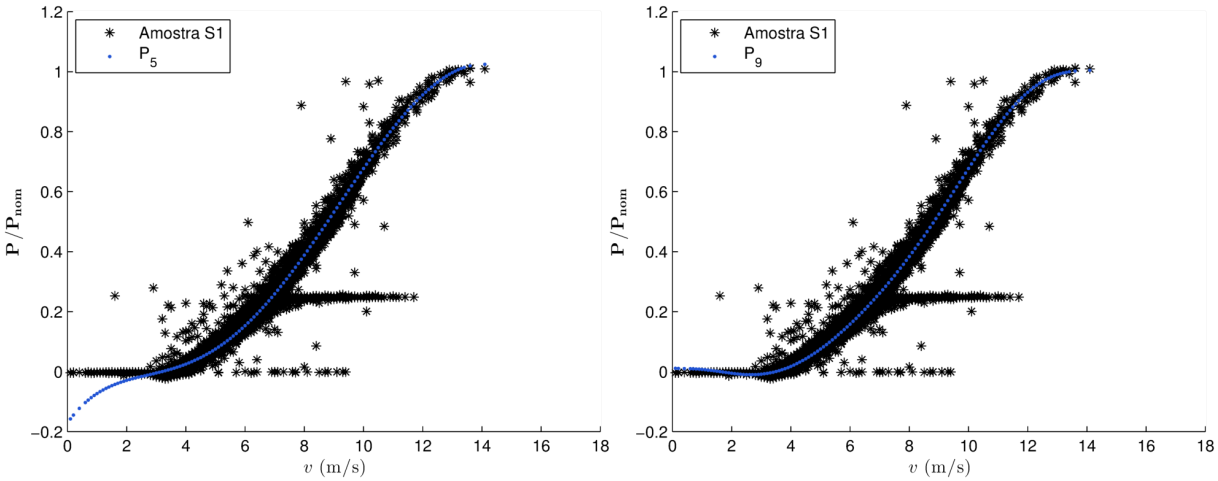
(b) Modelo NN+IM



Fonte – Próprio autor.

Figura 23 – CP polinomial grau 5 e grau 9 para o caso 1.

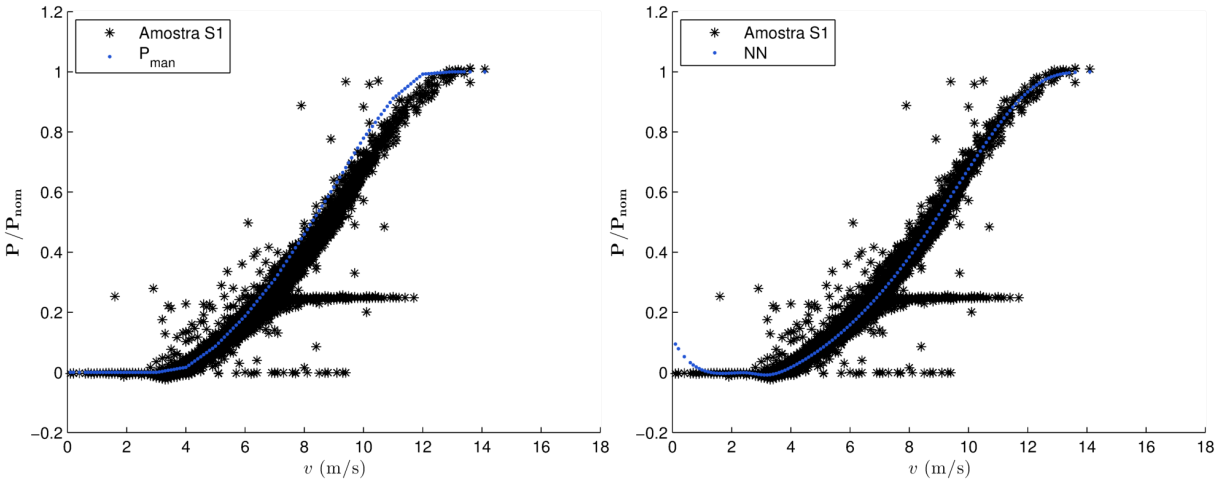
(a) Modelo P_5 (b) Modelo P_9



Fonte – Próprio autor.

Figura 24 – CP do fabricante e NN para o caso 1.

(a) Modelo do fabricante (b) Modelo NN

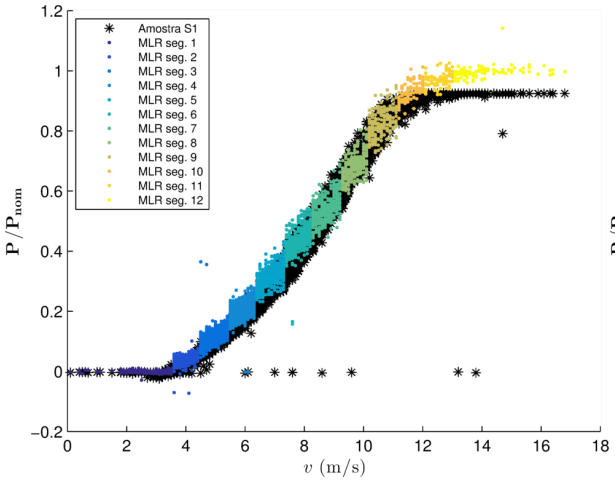


Fonte – Próprio autor.

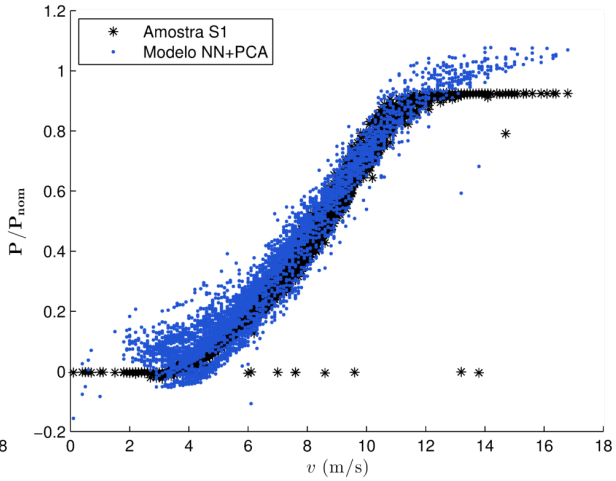
B.2 Caso 2

Figura 25 – Distorções da CP modelada utilizando dados provenientes da PCA para o caso 2.

(a) Modelo LS+PCA



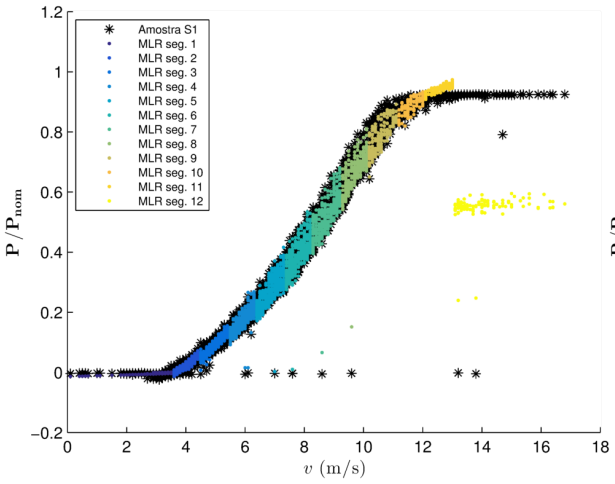
(b) Modelo NN+PCA



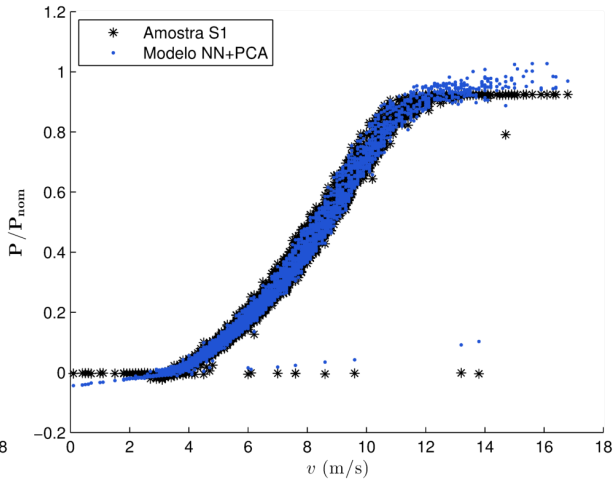
Fonte – Próprio autor.

Figura 26 – Distorções da CP modelada utilizando dados provenientes da IM para o caso 2.

(a) Modelo LS+IM



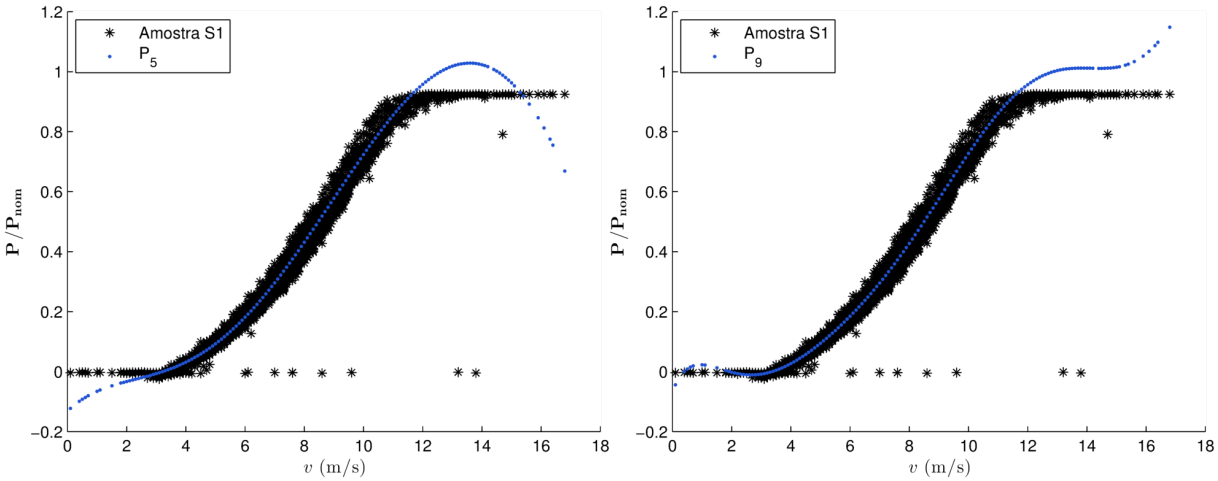
(b) Modelo NN+IM



Fonte – Próprio autor.

Figura 27 – CP polinomial grau 5 e grau 9 para o caso 2.

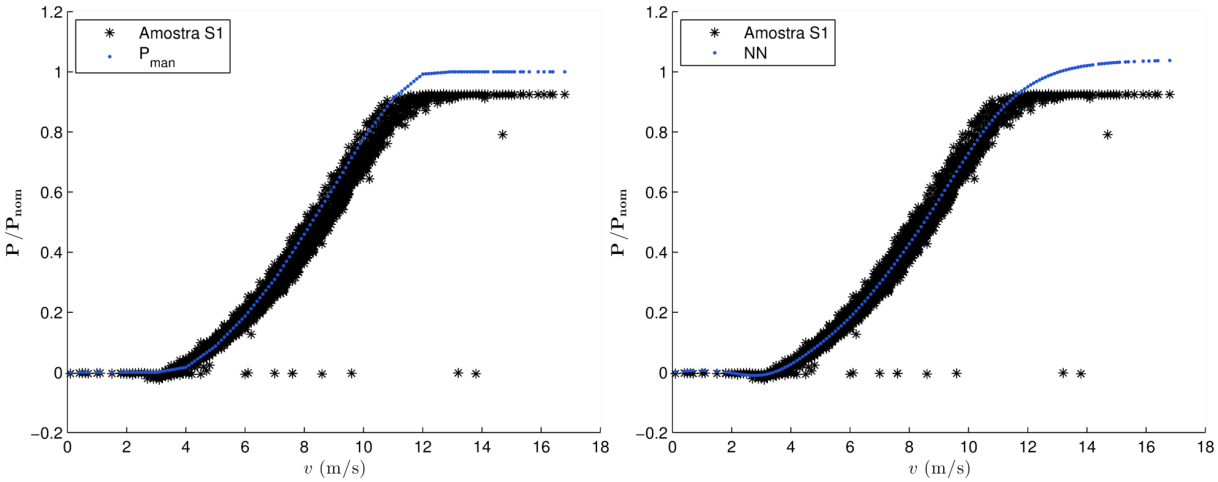
(a) Modelo P_5 (b) Modelo P_9



Fonte – Próprio autor.

Figura 28 – CP do fabricante e NN para o caso 2.

(a) Modelo do fabricante (b) Modelo NN



Fonte – Próprio autor.