

UNIVERSIDADE FEDERAL DE ALFENAS

LUÍS GUSTAVO GUTIERREZ GEBIN

**ABORDAGENS INTELIGENTES PARA ESTIMAR A PRODUÇÃO DE ENERGIA
EÓLICA**

Alfenas/MG

2018

LUIS GUSTAVO GUTIERREZ GEBIN

**ABORDAGENS INTELIGENTES PARA ESTIMAR A PRODUÇÃO DE ENERGIA
EÓLICA**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística Aplicada e Biometria, área de concentração em Estatística Aplicada e Biometria da Universidade Federal de Alfenas, MG, como parte dos requisitos para a obtenção do título de Mestre. Linha de Pesquisa: Modelagem Estatística e Estatística Computacional.
Orientador: Prof. Dr. Ricardo Menezes Salgado.
Coorientador: Prof. Dr. Denismar Alves Nogueira.

Alfenas/MG

2018

Dados Internacionais de Catalogação-na-Publicação (CIP)
Sistema de Bibliotecas da Universidade Federal de Alfenas

G293a Gebin, Luis Gustavo Gutierrez.
Abordagens inteligentes para estimar a produção de energia eólica / Luis Gustavo Gutierrez Gebin -- Alfenas/MG, 2018.
103 f.: il. -

Orientador: Ricardo Menezes Salgado..
Dissertação (Mestrado em Estatística Aplicada e Biometria) -
Universidade Federal de Alfenas, 2018.
Bibliografia.

1. Energia Eólica. 2. Modelos Estatísticos. I. Salgado, Ricardo Menezes.
II. Título.

CDD-519.5



MINISTÉRIO DA EDUCAÇÃO
Universidade Federal de Alfenas / UNIFAL-MG
Programa de Pós-graduação em Estatística Aplicada e Biometria

Rua Gabriel Monteiro da Silva, 700. Alfenas - MG CEP 37130-000
Fone: (35) 3299-1392 (Secretaria) (35) 3299-1121 (Coordenação)
<https://www.unifal-mg.edu.br/ppgeab/>



LUÍS GUSTAVO GUTIERREZ GEBIN

“ABORDAGENS INTELIGENTES PARA ESTIMAR A PRODUÇÃO DE ENERGIA
EÓLICA”

A Banca Examinadora, abaixo assinada, aprova a
Dissertação apresentada como parte dos requisitos para
a obtenção do título de Mestre em Estatística Aplicada
e Biometria pela Universidade Federal de Alfenas.
Área de Concentração: Estatística Aplicada e
Biometria

Aprovado em: 18 de junho de 2018.

Prof. Dr. Ricardo Menezes Salgado
Instituição: UNIFAL-MG

Assinatura: Ricardo Menezes Salgado

Prof. Dr. Flávio Barbieri Gonzaga
Instituição: UNIFAL-MG

Assinatura: Flávio Barbieri Gonzaga

Prof. Dr. Takaaki Ohishi
Instituição: UNICAMP

Assinatura: Takaaki Ohishi

*Dedico a todos que me apoiaram,
não é uma vitória só minha.*

AGRADECIMENTOS

Agradeço ao Laboratório de Inteligência Computacional (LInC) e a UNIFAL pela disponibilização dos recursos computacionais necessários para a execução dos experimentos e elaboração desta dissertação. A Dr. Dalia Patino-Echeverri da Duke University pelo compartilhamento de ideias e ajuda em relação aos métodos aplicados no trabalho. A CAPES pelo apoio financeiro. Ao meu orientador, Dr Ricardo Menezes Salgado, e ao meu orientador, Dr. Denismar Alves Nogueira, pela paciência em cada orientação.

RESUMO

Uma das maiores preocupações do século XXI é conciliar o desenvolvimento econômico e social com a preservação ambiental. Assim, o setor de energia torna-se foco de estudo, visto que grande parte das emissões de gases poluentes são advindas da geração elétrica. Como consequência, fontes renováveis e limpas de energia, que não causam malefícios para a natureza, como a produção de energia eólica, ganharam destaque no Brasil e no mundo. Porém, como a produção de energia eólica é uma energia obtida a partir do vento, o homem não possui total controle sobre sua geração, o que torna desejável que se tenha certa confiança no que concerne ao seu potencial elétrico. O que justifica o desenvolvimento de métodos eficientes para previsão destes dados eólicos, pois obtendo confiança quanto a estes modelos de previsão, pode-se expandir, ainda mais, a eficiência e potencial energético da produção de energia eólica brasileira, além de auxiliar as empresas no que se refere a distribuição da energia. Além disso, pode-se ter mais segurança em diminuir a complementação térmica, visto que as turbinas termoelétricas são usadas no Brasil, em geral, em períodos sazonais, quando a produção hidroelétrica está em baixa. Todavia, a previsão dos dados eólicos não é tarefa trivial, devido às diversas variáveis exógenas que podem afetar na produção. Visto isso, o trabalho tem como objetivo fazer a predição de uma semana em escala horária da produção de energia eólica nas quatro estações do ano com os modelos inteligentes (*XGBoost* e *Random Forest*) e com um modelo estatístico tradicional (*ARIMA*), para posteriormente aplicar nos modelos inteligentes um algoritmo de seleção de variáveis, além de propor um novo modelo baseado em *ensemble* com os modelos individuais. A partir dos resultados obtidos, pode-se perceber que o *Random Forest* foi o modelo mais beneficiado pela seleção, enquanto que o *XGBoost*, mesmo sem a seleção, conseguiu ter um desempenho interessante, visto o seu método de aproximação. O *ARIMA*, mesmo sem ter se ajustado perfeitamente, obteve resultado um pouco inferior aos modelos inteligentes. Quanto ao *ensemble* proposto, percebe-se que o mesmo foi superior aos modelos desenvolvidos individualmente, principalmente em relação ao *RMSE*. Vale ressaltar que o *ensemble* conseguiu melhorar na predição, principalmente, nos dados extremos, em que todos os modelos individuais superestimavam.

Palavras-chave: Energia eólica; Modelos estatísticos; Modelos inteligentes; Seleção de variáveis.

ABSTRACT

One of the major concerns of the twentieth century is to reconcile economic and social development with environmental preservation. Thus, the energy sector becomes the focus of study, since the part of emission of power gases is derived from the electric generation. As a consequence, renewable and clean energy sources, which do not cause harm to nature, such as a wind energy production, have gained prominence in Brazil and the world. Once again, the production of wind energy is an energy that can be taken out of the wind, man does not have total control over his generation, which makes it desirable that there is some confidence as to his electric potential. The justification for the development of these forecasting methods could be further extended by the energy potential of production and efficient in other years in terms of energy distribution. Besides that, you can have more security to decrease the thermal complementation, because the thermoelectric turbines are used in Brazil in seasonal periods, when hydroelectric production is low. However, an analysis of wind data is not a trivial task due to the existence of exogenous variables that can affect production. Seen this, the work aims to make the prediction of a week on an hourly scale of wind power production in the four seasons of the year with the intelligent models (Random Forest and XGBoost) and a ARIMA model, after that, apply in intelligent models a selection features, besides proposing a new model based on ensemble. From the results obtained, it can be seen that the Random Forest was the model most benefited by the selection feature, while the XGBoost, even without a selection, managed to perform interestingly, given its approach method. The ARIMA model, even without perfect fit, is a bit inferior to the smart models. As for the ensemble model, it is perceived that it was superior to the most disadvantageous models, especially in relation to RMSE. It is worth emphasizing that the ensemble model can improve in the prediction, mainly, in the extreme data, in which all the individual models overestimated.

Key-words: Wind power; Statistical models; Machine Learning; Feature Selection.

LISTA DE TABELAS

Tabela 1 –	Escalas de tempo e espaço dos movimentos atmosféricos.	23
Tabela 2 –	Os 10 maiores estados produtores de energia eólica - 2017 (jan-jun). .	40
Tabela 3 –	Produção de energia eólica em MW do primeiro semestre de 2011 e 2016.	57
Tabela 4 –	Produção de energia eólica em MW do segundo semestre de 2011 a 2016.	58
Tabela 5 –	Análise descritiva das semanas a serem preditas.	70
Tabela 6 –	Parâmetros para as construções dos modelos <i>Random Forest</i>	71
Tabela 7 –	Tempo de execução médio dos modelos <i>Random Forest</i> em segundos para a predição de uma hora.	72
Tabela 8 –	Parâmetros para as construções dos modelos <i>XGBoost</i> 1.	72
Tabela 9 –	Parâmetros para as construções dos modelos <i>XGBoost</i> 2.	73
Tabela 10 –	Tempo de execução médio dos modelos <i>XGBoost</i> em segundos para a predição de uma hora.	73
Tabela 11 –	<i>MAPE</i> das 168 horas de predições na primeira estação.	80
Tabela 12 –	<i>RMSE</i> das 168 horas de predições na primeira estação.	81
Tabela 13 –	<i>MAPE</i> das 168 horas de predições na segunda estação.	85
Tabela 14 –	<i>RMSE</i> das 168 horas de predições na segunda estação.	85
Tabela 15 –	<i>MAPE</i> das 168 horas de predições na terceira estação.	90
Tabela 16 –	<i>RMSE</i> das 168 horas de predição na terceira estação.	90
Tabela 17 –	<i>MAPE</i> das 168 horas de predições na quarta estação.	95
Tabela 18 –	<i>RMSE</i> das 168 predições na quarta estação.	95

LISTA DE FIGURAS

Figura 1 –	Usina hidrelétrica de Marimbondo.	19
Figura 2 –	Esquema de funcionamento da energia hidroelétrica.	20
Figura 3 –	Classificação do gás natural em sua origem.	22
Figura 4 –	Parque Eólica no mar no Reino Unido.	25
Figura 5 –	Turbinas eólicas de eixo horizontal e vertical.	26
Figura 6 –	Turbina Horizontal.	27
Figura 7 –	Curva de potência de uma turbina eólica.	28
Figura 8 –	Primeiro moinho de vento para geração de energia elétrica construída por Charles F. Brush.	29
Figura 9 –	Turbina Eólica Balaclava.	31
Figura 10 –	Máquina Smith-Putnam.	32
Figura 11 –	Turbina Eólica alemã moderna entre 1950 e 1960.	33
Figura 12 –	Participação europeia na produção eólica total até 2009.	34
Figura 13 –	Custo da Geração eólica em comparação com a geração a carvão e gás.	35
Figura 14 –	Parque eólico de Osório, RS.	37
Figura 15 –	Evolução mundial de energia eólica.	39
Figura 16 –	Energia eólica no mundo.	39
Figura 17 –	Fluxograma da atuação do ONS.	42
Figura 18 –	Ilustração de uma árvore de classificação/regressão.	48
Figura 19 –	Produção total do primeiro semestre de 2011 e 2016.	58
Figura 20 –	Produção total do segundo semestre de 2011 e 2016.	59
Figura 21 –	Produção total em MW de 2011 a 2017.	60
Figura 22 –	Fluxograma da construção dos modelos individuais.	62
Figura 23 –	<i>Ensemble</i> proposto.	63
Figura 24 –	Fluxograma do <i>Ensemble</i> proposto.	64
Figura 25 –	Os dados referentes a primeira semana alvo para a predição.	67
Figura 26 –	Os dados referentes a segunda semana alvo para a predição.	68
Figura 27 –	Os dados referentes a terceira semana alvo para a predição.	69
Figura 28 –	Os dados referentes a quarta semana alvo para a predição.	70
Figura 29 –	VarImportance dos dados totais com as 336 covariáveis.	75

Figura 30 – MAPE dos dias do <i>Random Forest</i> na primeira estação.	76
Figura 31 – MAPE dos dias do <i>XGBoost</i> para a primeira estação.	77
Figura 32 – Predição dos modelos <i>XGBoost</i> na primeira estação.	78
Figura 33 – Predição dos modelos <i>Random Forest</i> primeira estação.	78
Figura 34 – Predição do modelo <i>ARIMA(1,1,1)</i> na primeira estação.	79
Figura 35 – Predição do modelo <i>Ensemble XG</i> na primeira estação.	80
Figura 36 – MAPE dos dias do <i>Random Forest</i> na segunda estação.	81
Figura 37 – MAPE dos dias do <i>XGBoost</i> na segunda estação.	82
Figura 38 – Predição dos modelos <i>XGBoost</i> na segunda estação.	83
Figura 39 – Predição dos modelos <i>Random Forest</i> na segunda estação.	83
Figura 40 – Predição do <i>ARIMA(1,1,1)</i> na segunda estação.	84
Figura 41 – Predição do <i>ensemble XG</i> na segunda estação.	84
Figura 42 – MAPE dos dias do <i>Random Forest</i> na terceira estação.	86
Figura 43 – MAPE dos dias do <i>XGBoost</i> na terceira estação.	87
Figura 44 – Predição dos modelos <i>XGBoost</i> na terceira estação.	88
Figura 45 – Predição dos modelos <i>Random Forest</i> na terceira estação.	88
Figura 46 – Predição do modelo <i>ARIMA(1,1,1)</i> na terceira estação.	89
Figura 47 – Predição do modelo <i>ensemble</i> na terceira estação.	89
Figura 48 – MAPE dos dias do <i>Random Forest</i> na quarta estação.	91
Figura 49 – MAPE dos dias do <i>XGBoost</i> na quarta estação.	92
Figura 50 – Predição dos modelos <i>XGBoost</i> na quarta estação.	93
Figura 51 – Predição dos modelos <i>Random Forest</i> na quarta estação.	93
Figura 52 – Predição do modelo <i>ARIMA(1,1,1)</i> na quarta estação.	94
Figura 53 – Predição do modelo <i>Ensemble XG</i> na quarta estação.	94

SUMÁRIO

1	INTRODUÇÃO	13
1.1	OBJETIVOS E JUSTIFICATIVA	15
1.1.1	Objetivo Primário	15
1.1.2	Objetivos Secundários	15
1.1.3	Justificativa	15
1.2	ESTRUTURA DO TRABALHO	16
2	REFERENCIAL TEÓRICO	17
2.1	ENERGIA ELÉTRICA	17
2.1.1	Formas de geração elétrica e geração elétrica no Brasil	18
2.1.1.1	<i>Geração de energia limpa e renovável</i>	18
2.1.1.2	<i>Geração de energia suja e não renovável</i>	21
2.2	ENERGIA EÓLICA	23
2.2.1	Vento	23
2.2.2	Fator de capacidade	24
2.2.3	Geradores, turbinas e sistema de transmissão/distribuição	25
2.2.4	Histórico e produção de energia eólica no mundo	29
2.2.5	Produção de energia eólica no Brasil- História e dados atuais	37
2.2.6	Órgãos responsáveis pela energia elétrica no Brasil	40
2.3	MODELOS DE PREVISÃO PARA PRODUÇÃO DE ENERGIA EÓLICA	44
2.3.1	Estatísticos tradicionais- modelos Box-Jenkins (ARIMA)	45
2.3.2	Modelos inteligentes	47
2.3.2.1	<i>Classification and Regression Trees- CART</i>	47
2.3.2.2	<i>Extreme Gradient Boosting- XGBoost</i>	49
2.3.2.3	<i>Random Forest</i>	50
2.3.2.4	<i>Principal diferença entre o XGBoost e o Random Forest</i>	51
2.3.2.5	<i>Ensemble</i>	51
2.3.3	Seleção de variáveis	52
2.3.3.1	<i>Relevância</i>	52
2.3.3.2	<i>Importance- Random Forest</i>	53
2.4	REVISÃO BIBLIOGRÁFICA	54
3	MATERIAL E MÉTODOS	57
3.1	MATERIAL	57
3.2	DADOS	57
3.3	METODOLOGIA PROPOSTA	60
3.3.1	<i>Random Forest e XGBoost</i>	61
3.3.2	<i>ARIMA</i>	61
3.3.3	<i>Ensemble XG</i>	62
3.4	MÉTRICAS DE AVALIAÇÃO	65
4	RESULTADOS E DISCUSSÃO	66
4.1	ANÁLISE EXPLORATÓRIA	66
4.2	RESULTADOS DOS MODELOS	71
4.2.1	Parâmetros para a construção do modelo <i>Random Forest</i>	71
4.2.2	Parâmetros para a construção do modelo <i>XGBoost</i>	72
4.2.3	Variáveis selecionadas	74

4.2.4	Resultados para primeira estação	75
4.2.5	Resultados para segunda estação	81
4.2.6	Resultados para terceira estação	85
4.2.7	Resultados para quarta estação	90
4.2.8	Análise geral dos modelos	96
5	CONCLUSÕES E TRABALHOS FUTUROS	97

1 INTRODUÇÃO

Uma das grandes preocupações do século XXI é conciliar o desenvolvimento econômico e social, juntamente com a preservação ambiental. Assim, foco neste intuito se torna o setor de energia, visto que, segundo a Global Wind Energy Council - GWEC (2018), o mesmo é responsável por aproximadamente dois terços das emissões globais de gases poluentes no mundo. Vale destacar ainda, que assim como o aumento da população, existe uma tendência de aumento da demanda elétrica, o que causa certa preocupação com a preservação ambiental, além da social, visto a importância da eletricidade para as famílias e empresas. Estes fatores direcionaram um avanço científico e interesse maior para as fontes renováveis de energia, como por exemplo a energia eólica, para suprir futuros deficit sociais e degradações da natureza.

Como forma de produção elétrica, a mesma pode ser dividida em duas maneiras: suja ou limpa. Ou seja, a produção de energia suja ocorre quando a mesma resulta na liberação de resíduos, ou gases poluentes que prejudicam o meio ambiente e, em suma, são fontes não renováveis. A limpa, por outro lado, teoricamente, tal fato não ocorre, além de serem renováveis.

No que se concerne ao Brasil, a maior fonte de energia primária se refere a produção hidrelétrica, no qual é considerada limpa, renovável e de baixo custo de operação, porém, existem alguns pontos que merecem ser ressaltados. Por exemplo, um alto custo de implementação, impacto social e ambiental, devido à desapropriação de terras produtivas e desapropriação da população ribeirinha, além de ter seu potencial de produção extremamente influenciado pela seca. Devido esta baixa produção de energia elétrica em épocas de seca, se torna indispensável o uso de outras alternativas para suprir esta deficiência.

Segundo a Associação Brasileira de Energia Eólica, ABEE (2018), a energia hidroelétrica corresponde a 61% da matriz elétrica brasileira, seguido da energia termoelétrica com aproximadamente 9%. Esta energia termoelétrica é advinda da queima de combustíveis. O que acarreta a liberação de CO₂, no qual é considerado um gás poluente. Assim, cabe buscar recursos para diminuir esta complementação térmica.

A energia eólica, fonte de energia renovável e limpa, no qual abrange 8% da produção total, pode ser uma das saídas para tal complementação. Segundo a ABEE (2018), o Brasil, em 2018, tem capacidade instalada de 12,19 GW, porém ainda existe um grande potencial para desenvolvimento e expansão desta produção com cerca de 5,48 GW de capacidade já contratados e em construção para estarem prontos até 2020. Ou seja, a energia eólica já é responsável

por parte significativa no Brasil, além de prover bom potencial de crescimento. Além disso, o Brasil detém de ótimos ventos e de boa qualidade, principalmente no região nordeste. O fator de capacidade de vento no Brasil chega a ser superior a 50%, isto significa que 50% do dia se mantém uma produção relevante, enquanto a média mundial varia de 20% a 25%.

Apesar disso, por ser uma fonte de energia que provém do vento, e o homem não ter como produzi-lo, pode causar certa incerteza no que concerne a uma dependência de tal fonte na matriz elétrica brasileira. Assim, torna-se indispensável modelos de previsão para a produção da energia eólica para auxiliar, principalmente, o sistema elétrico brasileiro no planejamento de operação e da programação energética. Quanto a estes modelos de predição, pode se separar, segundo Lei et al. (2009), Foley et al. (2012), Holttinen, Miettinen e Sillanpää (2013) e Daraeepour e Echeverri (2014), em 4 metodologias: modelos meteorológicos, modelos estatísticos tradicionais (RUPPERT, 2011); (ELDALI et al., 2016); modelos inteligentes (LIU et al., 2010); (DARAEPOUR; ECHEVERRI, 2014); (YANG et al., 2015) e modelos híbridos (CHANG et al., 2016). Os modelos Meteorológicos são conhecidos como *Numerical weather prediction -NWP* e utilizam dados meteorológicos para realizar a predição, como a temperatura, pressão atmosférica, etc. Quanto aos modelos estatísticos tradicionais, os mais utilizados são os Box-Jenkinz, dentre eles os modelos *ARIMA* (RUPPERT, 2011) (ELDALI et al., 2016). Os modelos inteligentes como Redes Neurais Artificiais e *Support Vector Machine- SVM* são muito utilizados e ganharam muito destaque nos últimos anos (JURADO et al., 2013). Já os modelos Híbridos podem ser entendidos como combinações de modelos, também conhecido como *Ensemble*. O método *ensemble* consiste na junção de características de modelos construídos com a finalidade da criação de um modelo com melhor desempenho. Os algoritmos *XgBoost* e *Random Forest* se encaixam nesse perfil de modelo. Outro ponto necessário destacar, é acerca das diversas covariáveis nas quais podem afetar o modelo de previsão, como diversos dados climáticos. Como Daraeepour e Echeverri (2014) e Li, Wang e Goel (2015) afirmam, que, além de ter um modelo bem estruturado, também se torna indispensável a seleção das variáveis que serão inseridas no mesmo. Pois, a inserção de muitas covariáveis podem até atrapalhar a eficácia do modelo no que se refere a predição e deixar o algoritmo lento.

Com isso, o trabalho tem como objetivo fazer a predição de uma semana em escala horária da produção de energia eólica nas quatro estações do ano com os modelos inteligentes (*XGBoost* e *Random Forest*) e com um modelo estatístico tradicional (*ARIMA*), para posteriormente aplicar nos modelos inteligentes um algoritmo de seleção de variáveis, além de propor

um *ensemble* com os modelos individuais.

1.1 OBJETIVOS E JUSTIFICATIVA

Na Seção 1.1.1 será apresentado o objetivo primário da dissertação. Na Seção 1.1.2 serão abordados os objetivos específicos, e, por fim, nas Seções 1.1.3 e 1.2, serão apresentados as justificativas e estrutura do trabalho, respectivamente.

1.1.1 Objetivo Primário

O objetivo principal do trabalho é fazer a predição de uma semana, em escala horária, da produção de energia eólica para as 4 estações do ano.

1.1.2 Objetivos Secundários

- (a) Aplicar um método de seleção de variáveis nos modelos inteligentes;
- (b) Propor um *ensemble* com os modelos desenvolvidos individualmente;
- (c) Comparar os resultados dos modelos inteligentes com os modelos tradicionais, além do *ensemble*;
- (d) Relacionar os resultados ao longo das estações.

1.1.3 Justificativa

É sabido que uma das maiores preocupações do século XXI é acerca do bem-estar social e ambiental. A energia elétrica, e, as formas de como as mesmas são produzidas, entram nesta intersecção do bem-estar citada acima. Então, a solução para tal, é buscar formas de produção de energia que não denigrem o meio ambiente, ao mesmo tempo que apresentem formas eficientes

e constantes de energia.

Também é importante entender que erros na predição/planejamento de quanto se pode produzir podem causar um prejuízo enorme para o país, visto que podem levar a uma falta de energia.

A energia eólica é uma forma de produção de energia limpa e renovável, pois dependem da energia cinética do vento, porém, o homem não pode produzir o vento, o que pode o torna refém do mesmo. Além disso, é uma fonte que possui um potencial grande a ser explorado no Brasil devido ao bom fator de capacidade encontrado no país.

Outro ponto relevante é acerca das diversas variáveis exógenas que podem influenciar na produção do mesmo, dito isto, torna-se justificável que existam modelos preditores eficientes, além de métodos para seleção de variáveis que ajudem tais modelos, pois a inserção de muitas variáveis podem afetar negativamente na qualidade do mesmo. Além disso, conforme Ferreira (2012), é necessário que existam modelos que estimem a previsão diária de hora em hora, para permitir a aproximação do potencial energético a ser gerado e liberado na malha elétrica controlada pelo Operador Nacional de Sistema Elétrico (ONS).

1.2 ESTRUTURA DO TRABALHO

Para alcançar os objetivos propostos, o trabalho está estruturado como segue: na Seção 2 abordará acerca da produção energética brasileira, o histórico da produção de energia eólica no Brasil e no mundo, além de apresentar os modelos de previsão que são comumente usados para a predição da energia eólica, como também acerca dos métodos de seleção de variáveis. Na Seção 3 apresentará os dados e os modelos que serão utilizados no presente trabalho. Na Seção 4 discorrerá sobre os resultados e quais foram os parâmetros utilizados para cada modelo e, por fim, na Seção 5 serão apresentadas as conclusões do trabalho e sugestões para trabalhos futuros.

2 REFERENCIAL TEÓRICO

Na Seção 2.1 será apresentado acerca da importância da energia elétrica nos momentos atuais, além dos diferentes modos de geração de energia e suas relações. Na Seção 2.2 mostrará sobre o histórico e importância do vento, modo de produção e geradores. Quanto a Seção 2.3 serão apresentados os métodos e modelos mais comumente usados na literatura no que se refere a previsão de energia eólica, dando ênfase nos modelos utilizados no presente trabalho, além dos métodos de seleção de variáveis.

2.1 ENERGIA ELÉTRICA

A energia elétrica se tornou indispensável para o desenvolvimento econômico, social e industrial desde o século XX. É extremamente difícil para uma família ou empresa se desenvolver e sobreviver sem a referida energia, principalmente no que se concerne a um mercado competitivo. Em contrapartida, o modo no qual ela pode ser gerada se torna uma das preocupações deste século. Isto acontece devido a relação existente entre a produção de energia elétrica com impactos ambientais gerados por elas.

Neste contexto, assim como pontua a GWEC (2018), a produção elétrica é responsável por cerca de 2/3 da emissão dos gases CO₂ em todo planeta. Em contrapartida, a demanda elétrica, assim como a população em geral, tendem a crescer. O que gera certa preocupação e interesses voltados para as formas de produção de energia elétrica. Ou seja, neste contexto, além da preocupação ambiental, também existe a social, que contempla a abrangência dos setores elétricos.

Assim, para abordar melhor acerca do assunto, o trabalho apresentará na Seção 2.1.1 as formas de geração de energia elétrica, abordando as formas de geração limpa e suja. Posteriormente, discorrerá sobre os órgãos responsáveis pela energia elétrica no Brasil.

2.1.1 Formas de geração elétrica e geração elétrica no Brasil

Existem diversas formas de geração elétrica conhecidas e utilizadas nos dias atuais. Estas gerações podem ser divididas, *a priori*, em dois grupos: a suja, no qual denigrem o meio ambiente resultando a liberação de resíduos ou gases e não são renováveis, e a limpa, que se tornam tendência em um mundo com preocupações acerca da qualidade de vida e com o bem-estar ambiental, além de serem renováveis.

Estas formas de produção limpas e renováveis tendem a ter uma produção sustentável, porém, em alguns casos, não dependem, unicamente, da vontade do homem para a sua produção, visto a relação com as mudanças meteorológicas, e outras variáveis exógenas. Destacam-se as hidroelétricas, biomassa, solar, geotérmica, do mar, e advinda do ar. Na Seção 2.1.1.1 serão abordados mais especificamente sobre cada uma.

Quanto as fontes de energia suja, são formas de geração de energia elétrica não renováveis e não sustentáveis, além de causarem diversos impactos na natureza e, em alguns casos, trazerem malefícios, também, para as pessoas que convivem perto. Porém, é importante ressaltar, que o homem tem mais controle no que se refere a produção destas energias sujas. As mais utilizadas são o gás natural, nuclear e derivados de petróleo. Na Seção 2.1.1.2 serão abordados mais especificamente acerca de cada uma.

2.1.1.1 Geração de energia limpa e renovável

No que se refere na geração de energia limpa no Brasil, vale destacar a hidroelétrica (Figuras 1 e 2), principalmente em porcentual de participação na matriz elétrica nacional, tal fato se concretiza devido o imenso potencial hídrico existente. Ela não emite gás poluente, é uma fonte renovável e tem um baixo custo de operação (R\$ 153,70 por MWh), conforme GWEC (2018), porém, demanda de um alto custo de investimento inicial, um alto impacto ambiental, além do impacto social na desapropriação de terras ribeirinhas.

Figura 1 – Usina hidrelétrica de Marimbondo.

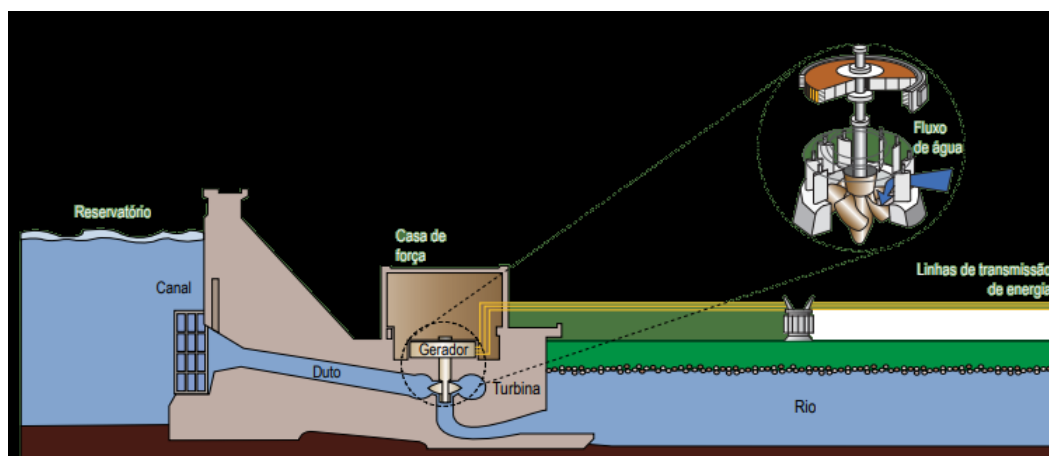


Fonte: Banco de imagens de Furnas.

Esta energia hidrelétrica é advinda do aproveitamento do fluxo das águas. Porém, além da necessidade da construção em grande escala, também é necessário, em alguns casos, o desvio do rio e a formação de reservatórios. Ou seja, torna-se necessária uma apropriação de terras, assim como uma construção e demanda recursos iniciais de mesmas escalas.

A produção hidroelétrica no Brasil, até 2001, tinha uma oferta quase que absoluta na participação elétrica nacional, ou seja, uma dependência altíssima de tal fonte. Porém, em 2001 o país passou por uma crise no setor elétrico, gerado, principalmente, pela falta de chuva e por planejamento inadequado dos responsáveis em questão, como afirma Tolmasquim (2000). Após esta crise, o país começou a se preocupar com a diversificação na matriz elétrica. Em 2018, segundo a ABEE (2018), as usinas hidrelétricas compõem cerca de 61% de toda produção elétrica brasileira.

Figura 2 – Esquema de funcionamento da energia hidroelétrica.



Fonte: (ANEEL, 2018).

Quanto a Biomassa, segundo a Agência Nacional de Energia Elétrica- ANEEL, do ponto de vista energético é toda matéria orgânica, seja de origem animal ou vegetal, que pode ser utilizada na produção de energia. Pode-se entender que a biomassa é uma forma indireta de energia solar, visto que a energia solar é convertida em energia química, através da fotossíntese, base dos processos biológicos dos seres vivos. Apesar de ter sua eficiência abaixo de outras formas de produção, principalmente em relação as formas de produção suja, a biomassa também possui suas vantagens, como seu aproveitamento de forma direta por intermédio da combustão de fornos e caldeiras (ANEEL, 2018).

A energia Solar é a energia advinda do sol, porém, pode-se entender que quase todas as fontes de energia (hidráulica, biomassa, eólica, combustíveis fósseis e energia dos oceanos) também são formas indiretas de energia solar (ANEEL, 2018).

Como energia gerada diretamente, pode se destacar os materiais termoelétrico e fotovoltaico. Segundo a ANEEL (2018), o termoelétrico se caracteriza quando a luz do sol é transformada em calor pela absorção das superfícies escuras. Enquanto que o fotovoltaico, os fótons contidos na luz solar são convertidos em energia elétrica, por intermédio das células solares.

Já a energia geotérmica é definida como a energia térmica proveniente do interior da Terra. Este método pode ser utilizado, principalmente, para aquecimento de água ou do ambiente residencial, por intermédio de bombas de calor ou *Ground-Source Heat Pump - GSHP* (ARBOIT et al., 2013). De acordo com Lund, Freeston e Boyd (2011), no Brasil foi identificado certo potencial para sistemas de alta temperatura geotérmica, porém, este recurso está limitado às ilhas atlânticas de Fernando de Noronha e Trindade.

Quanto a energia do mar, segundo Soerensen e Weinstein (2008), pode-se separar 5 formas distintas: A energia de ondas, energia da maré, energia de correntes, energia do gradiente de temperatura, energia do gradiente de salinidade. A energia das ondas é extraída por seu componente cinético através das turbinas. A energia corrente é a energia cinética extraída das correntes marinhas. A energia de correntes é a energia cinética extraída das correntes marinhas através de dispositivos modulares. A energia do gradiente de temperatura é advinda da diferença de temperatura entre as águas quentes superficiais e as águas geladas profundas. Quanto que a energia do gradiente de salinidade é a energia gerada a partir da energia osmótica da diferença de salinidade entre a água doce dos rios e a água salgada dos em estuários através de processos de osmose reversa e tecnologias de conversão associadas (SOERENSEN; WEINSTEIN, 2008); (FLEMING,).

Conforme Fleming (), estas energias não necessariamente teriam que ser usadas apenas para geração de eletricidade, como também para a geração direta de água potável ou para suprir necessidades térmicas, como o resfriamento.

No que se refere a energia eólica, que é gerada a partir da energia cinética do vento, foco do trabalho , será abordado na Seção 2.2.

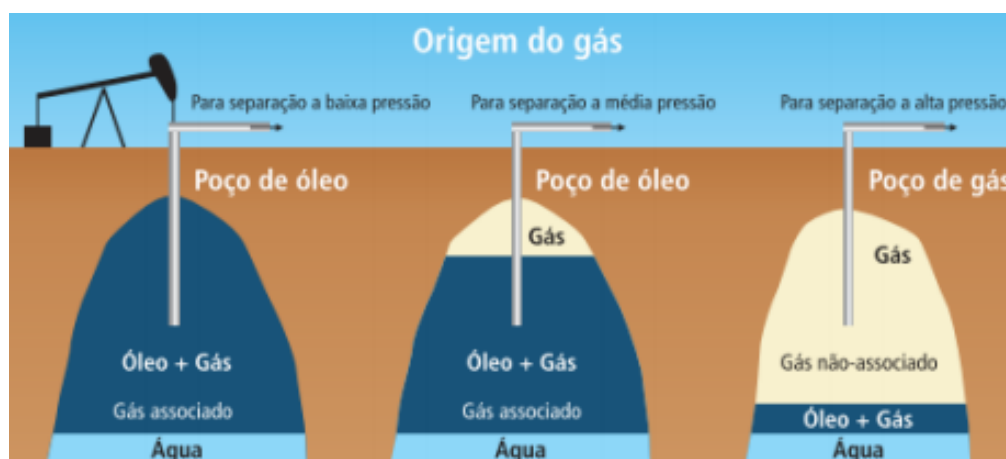
2.1.1.2 *Geração de energia suja e não renovável*

Segundo a ABRACE- Associação Brasileira de Grandes Consumidores Industriais de Energia e de Consumidores Livres (2012), o gás natural é um combustível fóssil que se encontra na natureza, e é formada por hidrocarbonetos com predominância do metano, e à temperatura ambiente e pressão atmosférica permanece em estado gasoso. Ele pode ser encontrado no subsolo em rochas porosas, isoladas do exterior por rochas impermeáveis. Suas reservas podem estar associadas a depósitos petrolíferos, mas não necessariamente a isso. Ela é resultado da degradação da matéria orgânica de forma anaeróbica oriunda de grandes quantidades de microorganismos que se acumularam, nas eras pré-históricas, nas águas litorâneas dos mares da época.

A geração de energia elétrica a partir de gás natural é feita pela queima do gás combustível em turbinas a gás (ANEEL, 2018). O gás natural também podem ser usado nas indústrias como no transporte, por exemplo. Pontos importantes no que se referem as características do gás natural são acerca dos baixos índices de emissão de poluentes, comparando aos outros combustíveis fósseis e a rápida dispersão em caso de vazamentos.

De acordo com Fioreze et al. (2013) o gás natural pode ser classificado de duas formas: gás associado e gás não associado, como visto na Figura 3. No Brasil, segundo a ABEE (2018), o gás natural, no que se refere a produção elétrica, tem participação de 9% na matriz energética brasileira.

Figura 3 – Classificação do gás natural em sua origem.



Fonte: (VIEIRA et al., 2005).

Quanto a geração energética nuclear é produzida a partir do urânio e outros elementos, e se dá a partir da fissão dos átomos de urânio. É uma das fontes de energia mais utilizadas no mundo. Porém, no Brasil ela compõem apenas 1% da matriz energética. Segundo Cardoso et al. (2001), a maior vantagem de uma central térmica nuclear é a grande quantidade de energia que pode ser gerada para pouco material usado.

Porém, o que gera preocupação para as produções nucleares, além de não serem renováveis, é acerca dos acidentes, que podem envolver muitas pessoas e geram malefícios que podem permanecer no local por certo tempo. Um dos casos mais famosos e recentes (2011) foi o de Fukushima, no Japão.

Por fim, segundo a ANEEL (2018), a geração elétrica que se dá dos derivados de petróleo geram a emissão de poluentes na atmosfera, principalmente os chamados gases de efeito estufa (GEE), dos quais os mais problemáticos são o dióxido de carbono (CO₂), o metano (CH₄) e o óxido nitroso (N₂O). Esses gases de efeito estufa vem sendo associados a mudanças climáticas e ao aquecimento global. Aproximadamente 2/3 das reservas de petróleo no mundo estão localizadas no Oriente Médio, assim, quando existem conflitos em áreas com alta concentração de petróleo, existe uma tendência de crise na oferta para com os outros países e consumidores, visto que o Oriente Médio não é um alto consumidor, e sim um grande exportador. No Brasil,

os derivados de petróleo ultrapassam os 8% no cenário energético total nacional.

2.2 ENERGIA EÓLICA

Neste sub-tópico serão dissertados acerca do vento, suas causas, formas, velocidades, além das turbinas eólicas, o histórico da energia eólica no mundo e no Brasil e os dados atuais da produção de energia eólica no Brasil nos tempos atuais.

2.2.1 Vento

Pode-se entender que o vento é a movimentação do ar provocado pelo aquecimento gerado pelo sol em partes diferentes da terra. A energia eólica, por sua vez, é causada pela energia cinética do vento. Assim, conforme afirma Custódio (2013), é importante estudar o comportamento espacial e temporal do vento para maximizar o seu aproveitamento energético. Lima (2016) também destaca algumas escalas acerca do vento que merecem certo estudo e foco, como suas causas, as escalas de seus movimentos, os seus tipos, força e sua potência.

Existem escalas do movimento do ar, que são faixas de deslocamentos atmosféricos decorrentes de movimentos aleatórios das moléculas individuais na circulação envolvendo a atmosfera inteira. Estes movimentos podem ser divididos em 4 escalas de comprimento: circulação geral, escala sinóptica, mesoescala e microescala, assim como pode ser visto e explicado na Tabela 1 (LIMA, 2016).

Tabela 1 – Escalas de tempo e espaço dos movimentos atmosféricos.

Escala	Tempo	Comprimento (Km)	Exemplos
Circulação geral	Semanas a anos	1000 a 40000	Ventos de movimento e correntes de vento.
Escala sinóptica	Dias a semanas	100 a 5000	Ciclones, anticiclones e furacão
Mesoescala	Minutos a dias	1 a 100	Tornados, brisas e temporais
Microescala	Segundos a minutos	<1	Turbulências e rajadas

Fonte: (CUSTÓDIO, 2013); (LIMA, 2016).

Existem forças das quais podem afetar o movimento do vento, e a força de Coriolis é uma delas (LIMA, 2016). A força Coriolis é a aceleração provocada pela rotação da terra, no

qual pode afetar o movimento, velocidade e direção do vento. Segundo Lima (2016) o efeito da força de Coriolis sobre o vento gera os movimentos tipicamente circulares, ou espirais, em torno dos centros de pressão, que tendem a provocar deslocamentos de massas de ar entre o Equador e os polos.

Quanto a circulação do vento, ou circulação atmosférica, ocorrem quando existe a movimentação do ar, gerado pelas diferenças de temperatura e pressão na atmosfera terrestre. Assim, considerando que os raios atingem a terra de forma não uniforme ao longo da extensão, percebem-se diferentes dinâmicas e movimentações do ar em um nível global, Lima (2016) destaca que, próximo ao Equador existe um cinturão contínuo de baixa pressão devido à maior entrada de calor do sol, assim, divide-se como segue:

- (a) Célula Tropical;
- (b) Célula Temperada;
- (c) Célula Polar.

As células tropicais produzem ventos de sudeste no Hemisfério Sul e ventos de nordeste no Hemisfério Norte, chamados de ventos alísios. Quanto as células temperadas, pode ser entendida como inversa, prevalecendo os ventos de noroeste e sudoeste, respectivamente, Hemisfério Sul e Norte. Esses ventos são chamados ocidentais e são separados por regiões de alta pressão chamadas de regiões de calmarias. A terceira célula é chamada de polar, com ventos do leste (LIMA, 2016). Porém, os ventos, assim como suas circulações estão muito ligadas ao efeito sazonal. Ou seja, suas movimentações e circulações são típicas de determinadas estações ou épocas do ano.

2.2.2 Fator de capacidade

Basicamente, o fator de capacidade serve para avaliar o potencial de energia eólica de alguma região, e pode ser entendido como o percentual de efetividade. Ou seja, a porcentagem de um aerogerador desta região consegue se manter operando, ao longo do dia, que pode ser entendida conforme a Equação 1.

$$FC = \frac{x}{Pn * 8760(Horas)} \quad (1)$$

No qual FC é o fator de capacidade em porcentagem, x é a energia produzida no ano, e P_n é potência nominal disponibilizada no sistema. Koto (2009) afirma que um FC entre 35 e 45 são considerados bons regimes de ventos.

Viterbo (2008) afirma, no que se refere ao fator de capacidade, que a produção eólica no mar (*offshore*, como pode ser vista na Figura 4) apresenta melhor desempenho. Além de vantagens como a área plana e a não ocupação de espaços em terra, e menor impacto sonoro (SILVA, 2015). Porém, demanda um maior investimento para instalação, transmissão, manutenção e com os equipamentos.

Figura 4 – Parque Eólica no mar no Reino Unido.



Fonte: (SILVA, 2015).

2.2.3 Geradores, turbinas e sistema de transmissão/distribuição

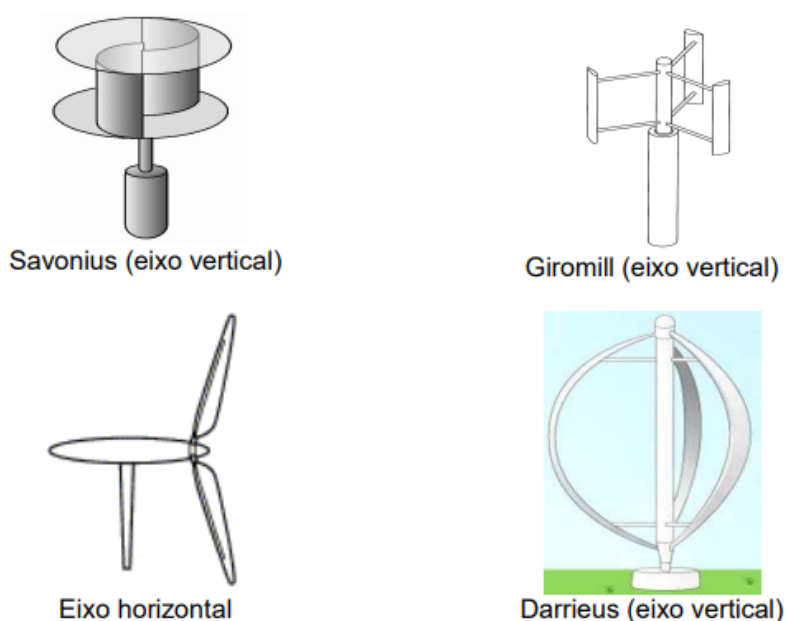
Os sistemas de transmissão e distribuição da aplicação para a tecnologia eólica de produção de energia elétrica podem ser divididas em dois modos: *onshore* e *offshore*. Basicamente, a *offshore*, como mostra Figura 4, é construída no mar.

Em outros casos, de acordo com Lima (2016), existem restrições importantes que devem ser considerados para a construção de aerogeradores mesmo em áreas que apresentem recursos

eólicos, como o tipo do terreno, a logística e infraestrutura pois podem influenciar no fator de capacidade. Além de dificultar na previsão dos dados, o que podem interferir na redistribuição de energia. Normalmente, evita-se a construção destes modelos em lugares muito ingrimes e montanhosos. Ou seja, os acessos devem ser amplos, devido ao tamanho dos materiais necessários, além disso, árvores grandes também podem influenciar os ventos locais. Outro fator importante a se considerar é acerca do acesso à rede elétrica.

Quanto aos modelos de geradores, turbinas, de energia eólica, pode-se separar em duas: vertical e horizontal, como pode ser visto na Figura 5.

Figura 5 – Turbinas eólicas de eixo horizontal e vertical.



Fonte: (LIMA, 2016).

A principal diferença acerca destas turbinas, é que as verticais não precisam de mecanismos direcionais, ou seja, sua produção independe da direção do vento. Além disso, as turbinas Darrieus e Savaonius podem ser instaladas no solo, ou seja, suas instalações são mais fáceis por não estarem em grandes alturas, o que facilita também a sua manutenção. A poluição sonora das referidas turbinas também são menores.

As turbinas de eixo horizontal precisam estar perpendiculares à direção do vento para maximizarem seu potencial energético, exemplo na Figura 6 e Figura 14. A altura das torres é importante para a geração eólica, além do diâmetro do rotor, pois sua potência gerada depende da mesma. Porém, devido a altura das turbinas, elas são mais eficientes do que os eixos verticais.

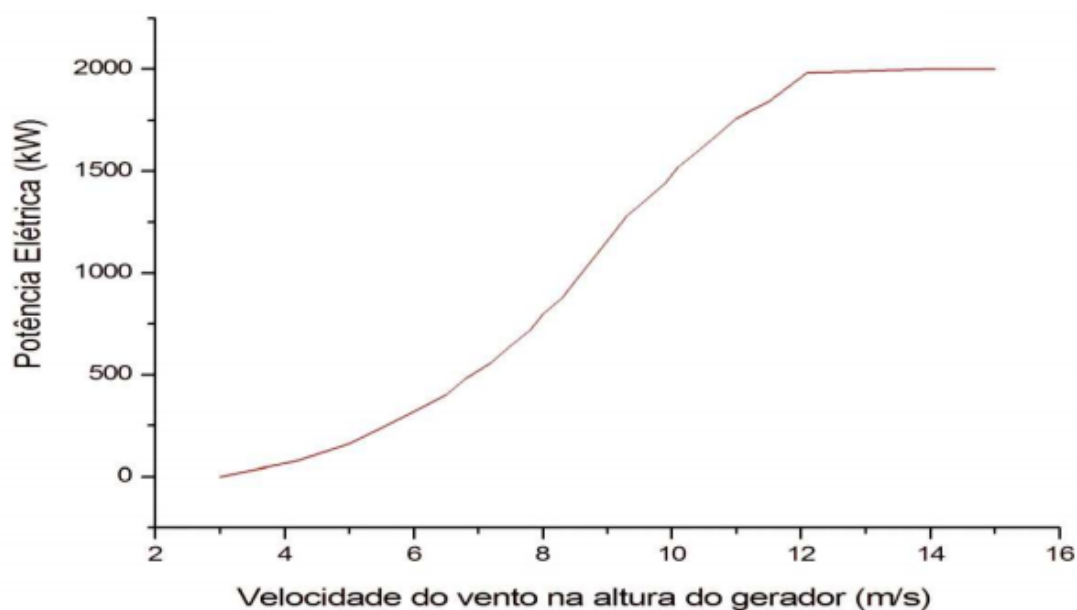
Figura 6 – Turbina Horizontal.



Fonte: (ANEEL, 2018).

A partir da Figura 7 disponibilizada por Ferreira (2012), percebe-se que a geração de energia elétrica se inicia a partir dos ventos com velocidade de pelo menos 3 m/s, assim como quase se estabiliza a partir dos 12 m/s. Além disso, Koto (2009) afirma que ventos acima de 25m/s a turbina fica comprometida devido ao material em que a mesma é constituída.

Figura 7 – Curva de potência de uma turbina eólica.



Fonte: LATEC, 2007.

Fonte: (FERREIRA, 2012).

A estrutura quanto a produção de energia elétrica por intermédio do vento baseia-se conforme Koto (2009): a energia cinética dos ventos é transformada em energia mecânica, para por fim transformá-la em elétrica. Os ventos impulsionam a hélice, que faz girar o rotor. Tal movimento é transferido para o conversor que por um processo de indução eletromagnética transforma a energia cinética dos ventos em energia elétrica.

Assim, segundo Koto (2009) existem diferentes classificações no que se referem a: número de pás das hélices, tipos de geradores, velocidade do rotor, pelo tipo de controle da potência mecânica e por tipo de ligação do gerador à rede elétrica.

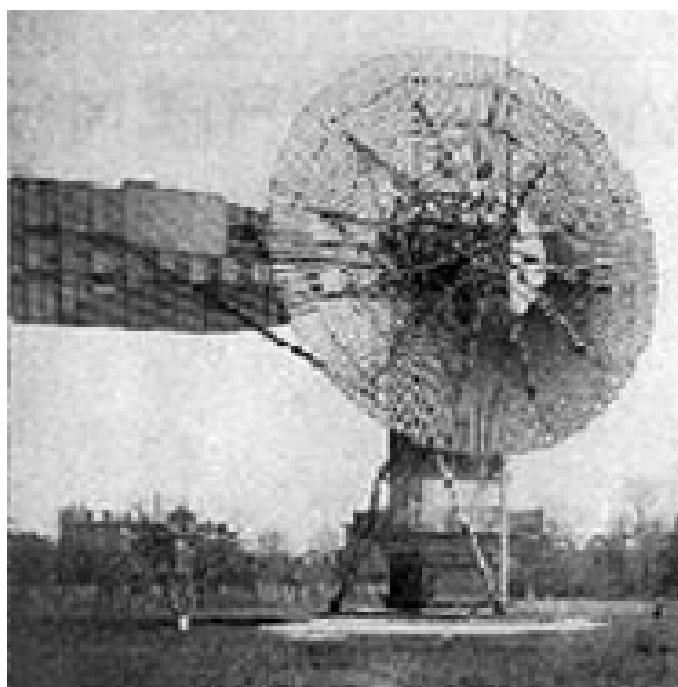
O número de hélices nas turbinas de eixos horizontais são quase sempre ímpares, devido a estabilidade (KOTO, 2009). Quanto aos tipos de geradores, destaca-se 3: geradores de corrente contínuas, geradores síncronos e geradores de indução. A velocidade do rotor pode ser constante ou variável. Os tipos de ligação do gerador à rede elétrica podem ser direto ou por conversor. E pelo tipo de controle da potência mecânica: por perda aerodinâmica ("*stall*") e por controle do ângulo de passo ("*pitch*").

2.2.4 Histórico e produção de energia eólica no mundo

O uso do vento para benefício humano não é um termo moderno, visto que tem uma tradição de mais de 2000 anos, segundo Kaldellis e Zafirakis (2011) existem registrados aparelhos de ventos mais simples em 200 A.C pelos persas. No século XIX, sistemas de vento evoluíram e, nos Estados Unidos, mais 6 milhões de pequenas maquinas que, usavam a energia do vento, eram utilizadas para bombear água entre 1850 e 1970 (DODGE, 2006).

Já no século XX, a energia advinda do vento passou a ser mais comumente utilizada para fornecer energia mecânica para bombear água ou moer grão, porém, com o inicio da industrialização moderna, o uso da energia eólica passou a ser substituídas por motores movidos a combustíveis fósseis ou com a energia hidroelétrica (ACKERMANN, 2005). No entanto, no ano de 1888, segundo Kaldellis e Zafirakis (2011), foi construído em Ohio pela empresa de Charles F. Brush, pela primeira vez, o uso do moinho de vento para transformar o vento em eletricidade (Figura 8). A máquina tinha 17 metros de diâmetro e produzia 12 kW.

Figura 8 – Primeiro moinho de vento para geração de energia elétrica construída por Charles F. Brush.



Fonte: (KALDELLIS; ZAFIRAKIS, 2011).

Em contrapartida, Dodge (2006) afirma que a primeira turbina para geração elétrica foi criada por Dane Poul Lacour em 1891 na Dinamarca. Esta tecnologia foi aperfeiçoada pelos dinamarqueses para suprir a energia elétrica durante a primeira e segunda guerra mundial. Já

em 1941-42, a empresa dinamarquesa F.L. Smidth b desenvolveram o que pode ser considerado a primeira turbina moderna, com algumas semelhanças das que são conhecidas no momento (DODGE, 2006) .

As primeiras turbinas eólicas para geração de eletricidade usavam pequenos propulsores modificados para acionar geradores de corrente contínua. Em meados de 1920, Parris-Dunn e Jabobs construíram geradores elétricos de 1 a 3 quilowatts no qual se faziam o uso nas áreas rurais de certas áreas americanas. Estes sistemas eram utilizados para iluminação e pequenos afazeres para as fazendas, porém, seu uso se estendeu para toda uma gama de aparelhos, como refrigeradores, freezers, etc.

Em contrapartida, entre os anos 1930 e 1940 estes sistemas foram se defasando. Os motivos foram o aumento da demanda de energia nas fazendas e a Grande Depressão, pois estimulou o governo federal dos EUA a incentivar as economias rurais estendendo a rede elétrica através dessas áreas. Porém, a energia eólica apenas ficou estagnada por certo tempo, e pode-se entender que apenas preparou o terreno para a possibilidade de um desenvolvimento futuro na região. Enquanto isso, outras regiões do mundo, principalmente a Europa, continuaram a se desenvolver.

Em 1931, na Rússia, pode ser observado, pela primeira vez na história, o uso da energia eólica para utilidade pública. A Turbina gerava 100 kW e operou por dois anos (Figura 9). As usinas eólicas experimentais posteriores nos Estados Unidos, Dinamarca, França, Alemanha e Grã-Bretanha durante o período de 1935 a 1970 mostraram que as turbinas eólicas de larga escala funcionariam, mas não resultariam de uma grande turbina eólica elétrica (THOMAS; SAVINO, 1973). (BRUHN; LORENSSON; SVENSSON, 2009)

Figura 9 – Turbina Eólica Balaclava.



Fonte: (THOMAS; SAVINO, 1973).

A maior turbina para geração de energia advinda do vento foi a máquina Smith-Putnam de 1,25 megawatt (Figura 10), instalada em Vermont em 1941. Esse desenho de eixo horizontal apresentava um rotor orientado para baixo da torre, de duas lâminas e 53,34 metros de diâmetro. Este gerador eólico foi o maior já construído com uma classificação de 1,25 megawatts em um 30 milhas por hora de vento uma torre de aproximadamente 33,53 metros (THOMAS; SAVINO, 1973).

Figura 10 – Máquina Smith-Putnam.



Fonte: (THOMAS; SAVINO, 1973).

Vale destacar que, após a Segunda Guerra Mundial, os preços dos derivados de petróleo aumentaram em demasia, assim, o mercado europeu volta a ver com bons olhos os investimentos no setor eólico.

Entre 1950 e 1960 a Alemanha obteve grande destaque no desenvolvimento da produção elétrica advinda do vento, devido a direção do Dr. U; Hutter. Máquinas de 10 kW a 100 kW foram desenvolvidas. As máquinas usavam lâminas leves de fibra de vidro com uma simples torre de tubo oco com fios de arame conforme visto na Figura 11. Vale destacar que esta turbina acrescentou muito no que se refere ao início de produção, visto que ele entregou produção nominal de 100 quilowatts a 18 milhas por hora. Enquanto que a maioria das outras máquinas estavam apenas começando a produção de energia neste cenário (THOMAS; SAVINO, 1973). Thomas e Savino (1973) também destacam que este esforço alemão representava, na época da produção do trabalho dos mesmos, a máquina mais moderna nesse tamanho.

Figura 11 – Turbina Eólica alemã moderna entre 1950 e 1960.



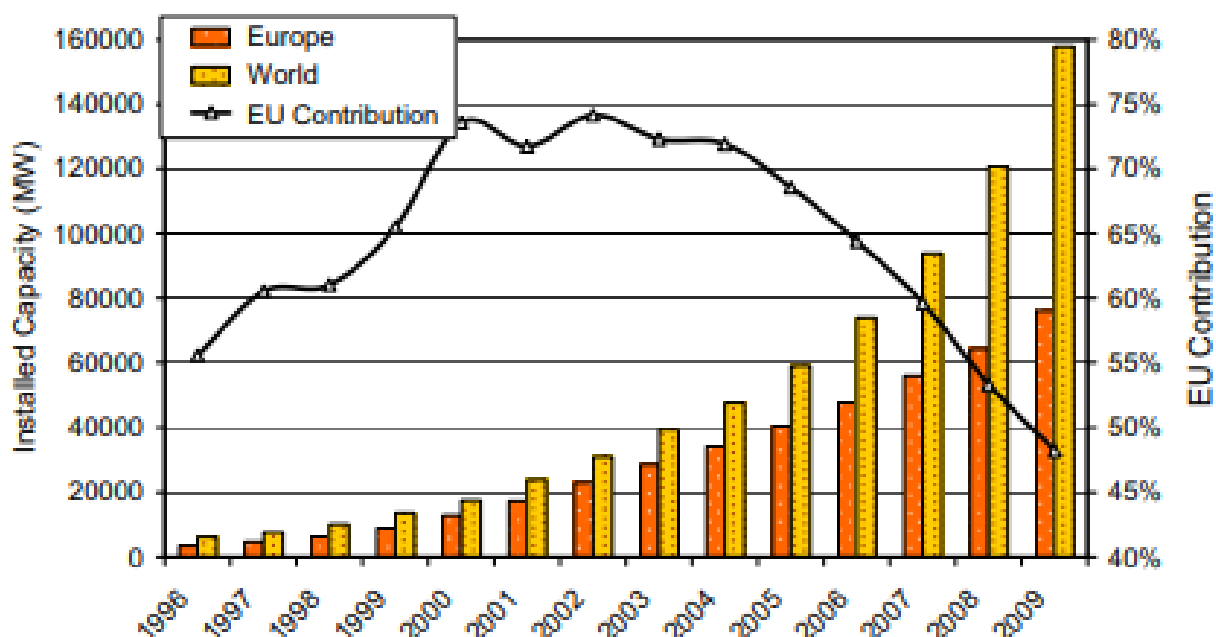
Fonte: (THOMAS; SAVINO, 1973).

No início dos anos 70, segundo Ackermann (2005) e Kaldellis e Zafirakis (2011), após o primeiro choque do preço do petróleo, interesses voltaram a ser direcionados para a energia eólica. Thomas e Savino (1973) já destacavam na época acerca do preço da produção da energia eólica. Os mesmos afirmavam que, apesar desta produção ainda ser mais cara, tinham futuro visto o aumento excessivo do preço dos combustíveis fósseis e, por consequente, a crise que viria.

Entre 1973 e 1986 o uso da energia eólica passou do uso doméstico e agrícola para a aplicação de parques eólicos interligados. Assim, segundo Kaldellis e Zafirakis (2011), surgiu a primeira produção em larga escala na Califórnia, no qual foram construídas mais de 16,000 máquinas.

Entre os anos 80 e 90, as instalações dos parques eólicos aumentaram constantemente no norte da Europa, onde, devido o alto custo da eletricidade, desenvolveu-se um mercado pequeno no que se refere a produção eólica, porém estável (KALDELLIS; ZAFIRAKIS, 2011). Após 1990, a maior parte destas atividades se concentraram no mercado europeu, no qual se tornaram linha de frente no cenário mundial, liderados pela Dinamarca, Alemanha, Espanha e Holanda (ACKERMANN; SÖDER, 2002).

Figura 12 – Participação europeia na produção eólica total até 2009.

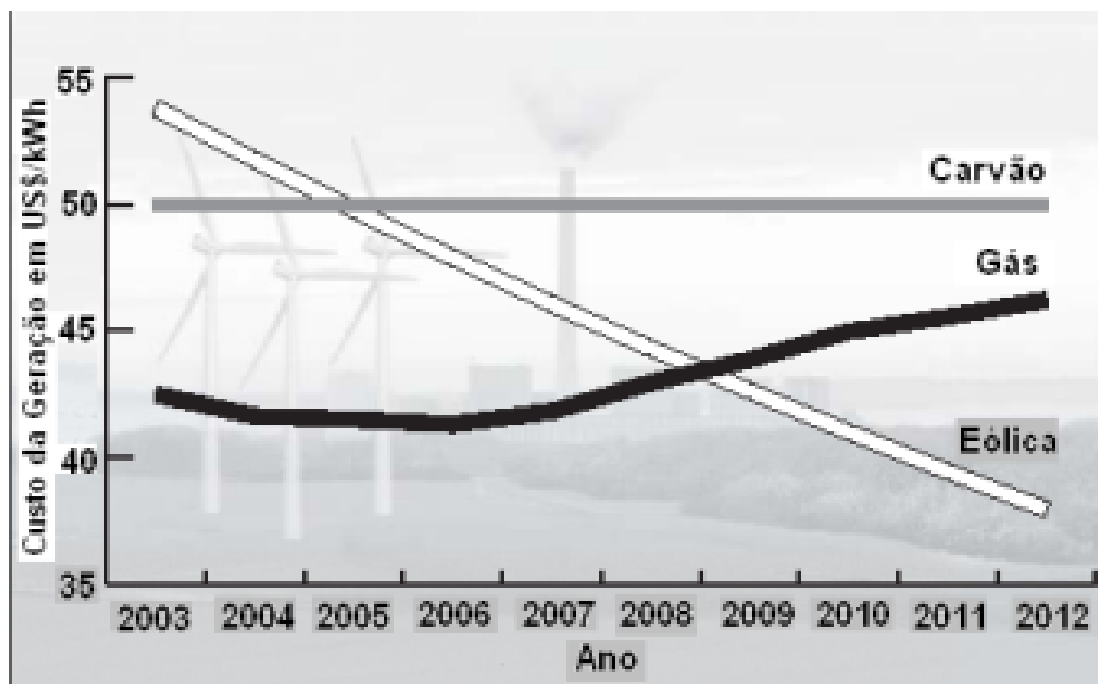


Fonte: (KALDELLIS; ZAFIRAKIS, 2011).

Assim, Kaldellis e Zafirakis (2011) afirma que, apesar do início de produção em massa ter seu início em 1990 no mercado europeu, no qual participavam do mercado nacional em cerca de 20% da produção total eólica no início, após os anos 2000 esta produção chegou a ser superior a 70% (Figura 12). Porém, devido a vários incentivos para produção eólica ao longo do mundo, principalmente na China e recomeço deste investimento pelos Estados Unidos, tal participação diminuiu muito, conforme visto na Figura 16, que mostra a participação no ano de 2016, que aponta a China como maior produtor de energia eólica em âmbito global.

Ainda de acordo com Kaldellis e Zafirakis (2011), um dos motivos para a expansão da produção eólica ao longo do mundo se dá pelo desenvolvimento científico das turbinas eólicas. E, conforme o preço de implantação destes materiais foram diminuindo, assim como a produção em si, a opção pela instalação foi aumentando, visto o baixo custo de operação da produção eólica comparada a outras produções elétricas. Como pode ser observado na Figura 13 (RÜNCOS et al., 2000)

Figura 13 – Custo da Geração eólica em comparação com a geração a carvão e gás.



Fonte: (RÚNCOS et al., 2000).

Outro ponto relevante para o investimento em fontes renováveis se deu a partir das preocupações com o meio ambiente. Porém, para entender melhor tal preocupação, é importante fazer uma análise histórica de como a mesma se desenvolveu ao longo da história.

O primeiro impacto que o homem gera na natureza, aparentemente, se dá pela agricultura. Devido ao aumento populacional, torna-se necessário um aumento na produção de alimentos, o que resulta no desmatamento para tal cultivo.

Assim, o próximo choque se deu com o desenvolvimento das manufaturas, o que precedeu a Revolução Industrial. Com o desenvolvimento das manufaturas já pode ser observado o início da dependência das fontes não renováveis (BORGES; TACHIBANA, 2005). Com a Revolução Industrial a exploração dos recursos naturais é cada vez mais relevante, assim como os resíduos e poluição por ela gerada.

Porém, as primeiras preocupações internacionais com o meio ambiente se sucederam, de acordo com Andrade, Tachizawa e Carvalho (2000), no século XX com a Conferência Científica da ONU sobre a Conservação e Utilização de Recursos em 1949, e com a Conferência sobre a Biosfera em 1968 realizada em Paris. Todavia, não obtiveram grande sucesso e destaque. Assim, pode-se separar em dois grandes marcos para o despertar da consciência ecológica no mundo: a publicação do Relatório Limites do Crescimento, elaborado pelo Clube de Roma e a

Conferência de Estocolmo, em 1972 (I CNUMAD), que teve foco a conscientização dos países acerca da importância da conservação ambiental como fator fundamental para a manutenção da espécie humana. Ressalta-se que a palavra-chave na conferência de Estocolmo foi poluição (BORGES; TACHIBANA, 2005).

Já para Borges e Tachibana (2005), pode-se destacar em três grandes marcos para a preocupação do planeta, além de 1949 e 1972: a Conferência das Nações Unidas para o Meio Ambiente e Desenvolvimento, realizada no Rio de Janeiro, em 1992, e o Protocolo de Kyoto.

A Conferência Rio-92 teve tanta importância, que Mebratu (1998) destaca que pela primeira vez em muitos países envolveu diversos segmentos da sociedade, levando assim o conceito de "desenvolvimento sustentável" para todas as partes do planeta.

Quanto o Protocolo de Kyoto, realizado e aprovado no ano de 1997, teve como o meta reduzir entre os anos de 2008 e 2012, em média 5,2% das emissões da atmosfera dos gases que provocam o efeito estufa, que são eles: dióxido de carbono, metano, óxido nitroso, hidrofluorcarbono, perfluorcarbono e o hexafluorcarbono de enxofre (BORGES; TACHIBANA, 2005).

De acordo com Koto (2009), o Protocolo de Kyoto apresentou o que ficou conhecido como Mecanismo de Desenvolvimento Limpo (MDL) para incentivação da diminuição dos gases de efeito estufa para os países. Assim, um dos indicadores do MDL era acerca do incentivo do aumento da porcentagem da produção elétrica sustentável, como a energia eólica.

Junqueira (2006) e Orsolon (2006) afirmam que o Brasil se destacou muito nos projetos de créditos de carbono registrados na ONU, no qual grande parte da redução dos gases de efeito estufa se sucedem devido aos projetos de geração elétrica, ou seja, grande parte da não emissão destes gases são devido aos projetos de geração elétrica sustentável. Orsolon (2006) ainda pontua que, para acrescentar créditos o projeto também precisa contribuir para o desenvolvimento sustentável.

Em suma, acerca da energia eólica, pode-se listar 3 principais ramos dos quais a mesma podem beneficiar: ambiental, econômico e social. Por ser uma energia produzida a partir dos ventos, não produz resíduos, assim, não polui o meio ambiente, sendo considerada uma energia limpa, além de renovável. Também auxilia na redução de emissão de CO₂, visto que podem substituir os métodos de queimas de combustíveis em certos níveis de aceitação, o que também contribui para com o acordo climático de Paris e no Protocolo de Kyoto. Quanto a parte financeira, segundo a Câmara de Comercialização de Energia Elétrica- CCEE e a ABEE (2018),

o preço médio atualizado da energia eólica contratada pelos leilões é de R\$172,60 por MWh, ficando atrás apenas da energia gerada pelas hidrelétricas com R\$ 153,70 por MWh. Porém, se considerar o impacto ambiental, além do impacto financeiro dos grandes projetos para as construções de novas hidrelétricas, a energia eólica se torna a mais rentável. Por fim, quanto a parte social, permite que o proprietário da terra continue com sua produção local, visto que a área efetiva varia entre 3% e 5%, ou seja, gera um auxílio financeiro para com este produtor rural, por meio de arrendamentos da terra para as instalações das torres eólicas, o que incentiva a produção rural para pequenos e médios produtores, como pode ser visto na Figura 14, ou seja, um pequeno produtor rural poderia utilizar o mesmo espaço dos parques, visto o espaço dos mesmos (ABEE, 2018).

Figura 14 – Parque eólico de Osório, RS.



Fonte: Prefeitura de Osório.

2.2.5 Produção de energia eólica no Brasil- História e dados atuais

No Brasil no ano de 2018, existem 518 usinas instaladas no qual produzem 13 GW de energia gerada, além de uma capacidade de 4,87 GW de produção em construção. Esta

¹<http://www.osorio.rs.gov.br>. Acesso em: 03 de abril de 2018.

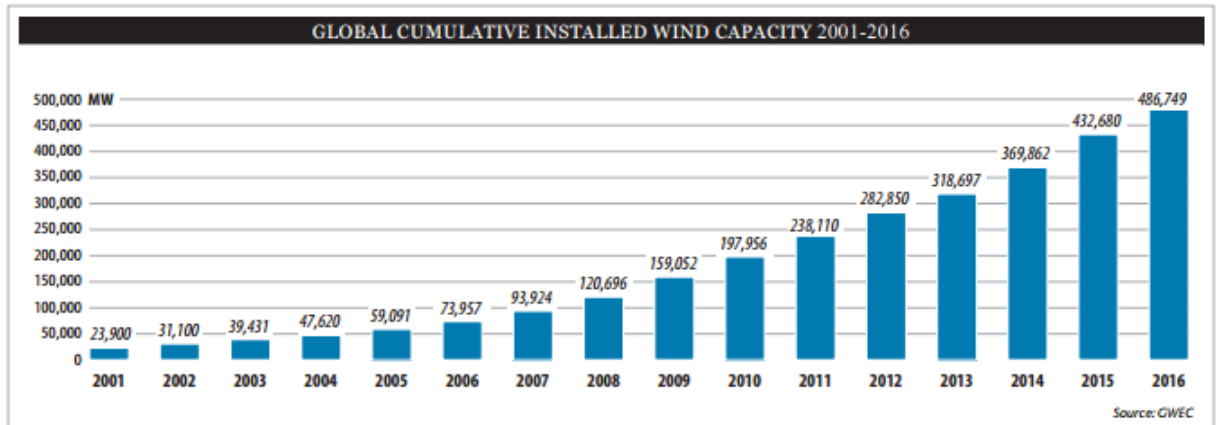
produção equivale a uma redução de 23.000.000 T/ano de CO₂, segundo a ABEE (2018), esta emissão de CO₂ evitada equivale a emissão de cerca de 11 milhões de automóveis.

A energia eólica no Brasil começou a sua operação comercial do primeiro aerogerador em 1992, no qual foi gerado a partir de uma parceria entre o Centro Brasileiro de Energia Eólica (CBEE) e a Companhia Energética de Pernambuco (CELPE), através de financiamento do instituto de Pesquisas dinamarquês Folkecenter. Esta turbina eólica foi a primeira a entrar em operação na América do Sul e esta localizada no arquipélago de Fernando de Noronha (ABEE, 2018).

A sua evolução, porém, ficou praticamente estagnada por dez anos. Contudo, após a crise energética de 2001 este cenário começou a mudar. Segundo Tolmasquim (2000), esta crise se deu, não apenas pela crise hídrica no qual o Brasil passou, mas também por falta de diversificação na matriz energética brasileira e descaso público. Ou seja, durante este período a demanda elétrica cresceu mais do que a produção e, juntamente com a crise de chuvas, fez com que o país passasse por um racionamento e aumento das contas de energia. Com esta crise, tornou-se necessário a diversificação do cenário elétrico brasileiro. Por conta disso, houve um certo incentivo na contratação de empreendimentos de geração de energia eólica no país. A partir de então, foi criado o Programa Emergencial de Energia Eólica - PROEÓLICA, na qual tinha como expectativa complementar a demanda elétrica nos comportamentos sazonais da produção hidroelétrica, ou seja, garantir a demanda elétrica nos períodos de baixa produção elétrica advinda da hidrelétrica, porém, não obteve tais resultados e o programa foi substituído pelo Programa de Incentivo às Fontes Alternativas de Energia Elétrica - PROINFA. Além do incentivo a outras fontes renováveis, o PROINFA também auxiliou na fixação da indústria de componentes e turbinas eólicas no Brasil (ABEE, 2018).

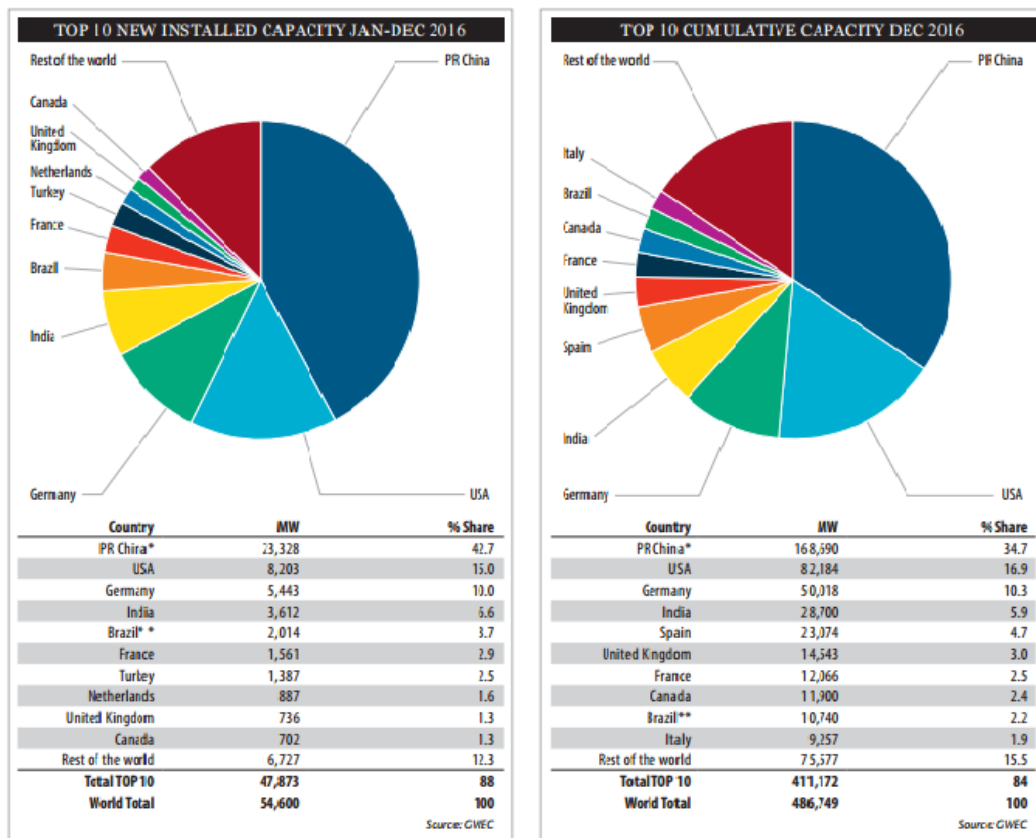
Quanto a produção, comparando em nível mundial, o Brasil ocupa a sétima posição em capacidade instalada. Porém, não foi apenas no Brasil que a produção de energia eólica se desenvolveu nos últimos anos. Na Figura 15 pode-se perceber uma tendência de crescimento. Já na Figura 16 percebe-se que o Brasil representa grande parte de tal crescimento, sendo um dos países com maior capacidade instalada no mundo.

Figura 15 – Evolução mundial de energia eólica.



Fonte: (ABEE, 2018).

Figura 16 – Energia eólica no mundo.



Fonte: (ABEE, 2018).

Além da produção crescente, o Brasil também é visto com bons olhos para investimento

nesta área. Segundo a Climatescope (2016), o Brasil está entre os melhores países para investimento de energia renovável. Tal estimativa se justifica devido o Brasil, principalmente no Nordeste, ter um fator de capacidade elevado, sendo superior a de 50% em algumas regiões, visto que a média mundial está entre 20% e 25%. A média nacional no ano em 2016 foi de 42% (ABEE, 2018).

Em 2017, a produção de energia eólica no Brasil continuou se desenvolvendo. Na região nordeste, por exemplo, de acordo com a ABEE a energia advinda do vento abasteceu cerca de 64% da energia consumida, com fator de capacidade de 76% no dia 14/09.

Tabela 2 – Os 10 maiores estados produtores de energia eólica - 2017 (jan-jun).

Posição	Estado	MW médios
1º	Rio Grande do Norte	1227,00
2º	Bahia	819,00
3º	Rio Grande do Sul	565,60
4º	Ceará	494,00
5º	Piauí	402,40
6º	Pernambuco	225,60
7º	Santa Catarina	27,50
8º	Paraíba	14,70
9º	Sergipe	7,40
10º	Rio de Janeiro	7,00

Fonte: Elaborada pelo autor.

Porém, é importante ressaltar que o homem não controla os ventos, ou seja, pode-se apenas transformar o vento em energia elétrica e não produzi-lo. O que pode gerar certa desconfiança em ter uma alta dependência da energia eólica na matriz energética brasileira.

2.2.6 Órgãos responsáveis pela energia elétrica no Brasil

O órgão responsável pela coordenação e controle das operações das instalações de geração e transmissão de energia elétrica no Sistema Interligado Nacional (SIN) e pelo planejamento de operações dos sistemas isolados do país é conhecido por Operador Nacional do Sistema Elétrico (ONS), no qual é fiscalizado e regulado pela Agência Nacional de Energia Elétrica (ANEEL).

O ONS foi criado em 26 de agosto de 1998, pela Lei nº 9.648, com as alterações introduzidas pela Lei nº 10.848/2004 e regulamentado pelo Decreto nº 5.081/2004. Ela é instituída

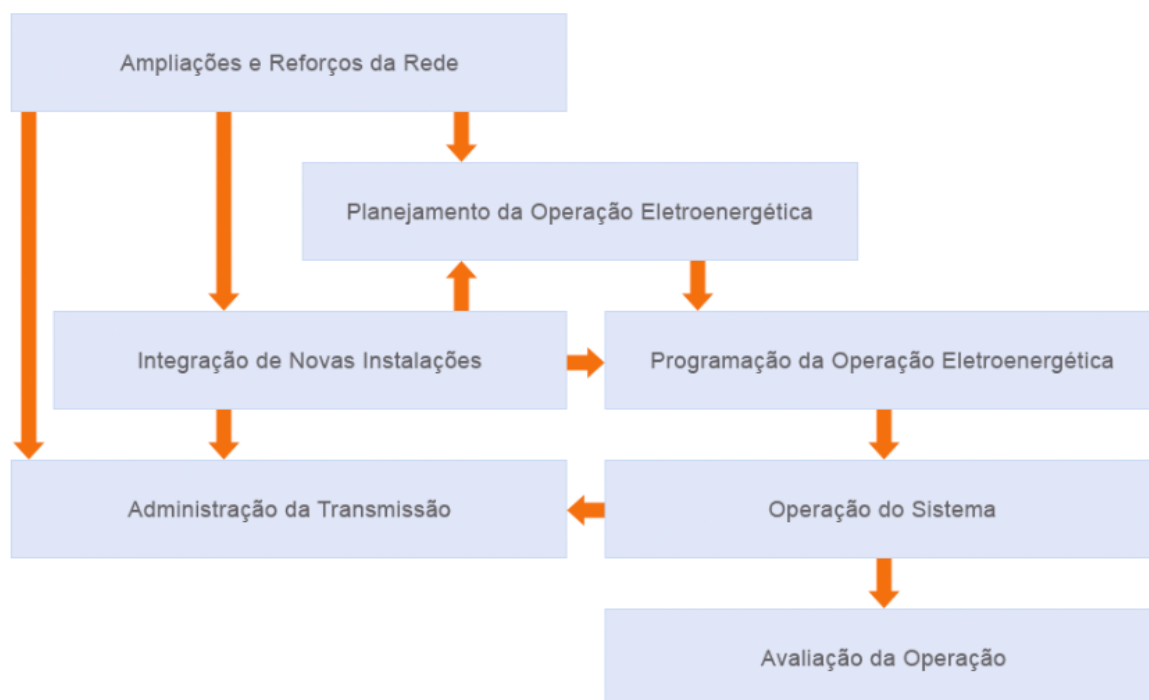
como uma pessoa jurídica de direito privado, sob forma de associação civil sem fins lucrativos.

Segundo o ONS (2018), o mesmo desenvolve estudos e ações exercidas sobre o sistema e seus agentes proprietários para gerenciar as diferentes fontes de energia e a rede de transmissão, de forma a garantir a segurança do suprimento contínuo em todo o país, no qual destacam-se 3 objetivos: Promover a otimização da operação do sistema eletroenergético, visando ao menor custo para o sistema, observados os padrões técnicos e os critérios de confiabilidade estabelecidos nos Procedimentos de Rede aprovados pela Aneel; garantir que todos os agentes do setor elétrico tenham acesso à rede de transmissão de forma não discriminatória; e contribuir, de acordo com a natureza de suas atividades, para que a expansão do SIN se faça ao menor custo e vise às melhores condições operacionais futuras.

Participam do ONS membros associados e membros participantes, compostos por empresas de geração, transmissão, distribuição, consumidores livres, importadores e exportadores de energia. Além da participação do Ministério de Minas e Energia (MME) e representantes dos Conselhos de Consumidores.

Para garantir os objetivos apresentados, o ONS se separa em segmentos, como pode ser percebido na Figura 17. Nela pode-se observar como se relacionam entre si estes diferentes segmentos, como: as ampliações e reforços de rede, as integrações de novas instalações, os planejamentos referentes a operação eletroenergética, Programação da Operação Eletroenergética, a Operação do Sistema, Avaliação da Operação e a Administração da Transmissão.

Figura 17 – Fluxograma da atuação do ONS.



Fonte: ONS.

Essas ampliações e reforços da rede são necessários para garantir as condições de segurança operativa do SIN. A partir dela é criada o Plano de ampliações e Reforços (PAR), no qual objetiva determinar as ampliações da Rede Básica e reforços em instalações de transmissoras e distribuidoras que causem impacto sistêmico. Tais instalações são consideradas no planejamento para expansão da geração e do sistema de transmissão, na avaliação do desempenho das interligações inter-regionais e internacionais, e no dimensionamento de compensação de potência reativa. Outro objetivo do PAR é avaliar as melhorias relacionadas à substituição de equipamentos por obsolescência, esgotamento de vida útil, falta de peças de reposição, risco de danos a instalações, desgastes prematuros ou restrições operativas intrínsecas, bem como em função da análise de superação de equipamentos (ONS, 2018).

No que se refere ao processo de Integração de Novas Instalações, a SIN envolve atividades sob responsabilidade do ONS, assim como dos agentes de operação responsáveis pelas instalações que fazem parte do SIN. Esse processo requer interações com praticamente todas as áreas do ONS, considerando a adequação física, operacional e legal da instalação. As leis nº 9.074/1995 e nº 9.648/1998 impõem o livre acesso, assim, asseguram o direito de qualquer

acessante a conectar-se e fazer uso do sistema elétrico (ONS, 2018).

Os planejamentos referentes a operação eletroenergética se dão a partir de estudos elétricos e energéticos de planejamento das operações dos quais realizam avaliações das condições futuras de suprimento e são baseadas nos critérios de otimização, recomposição, reserva e segurança operativa, além de análises da continuidade do suprimento eletroenergético. Segundo o ONS (2018), contemplam horizontes plurianual e anual, com detalhamento em base mensal, e têm como principais insumos cenários de carga e condições hidrológicas. Estes resultados abrangem estratégias de utilização dos recursos energéticos disponíveis e subsidiam a programação da operação eletroenergética e a pré-operação para a elaboração de despachos de geração que permitam utilizar esses recursos para atender ao mercado com segurança e minimizar o custo da operação.

Quanto a Programação da Operação Eletroenergética, a mesma é desenvolvida a partir da avaliação de curto prazo das condições de suprimento. Com isso, são estabelecidas diretrizes para a operação elétrica em âmbitos quadrimestrais e mensais, e definido o Programa Mensal de Operação (PMO), além de suas revisões semanais, com as diretrizes para a operação energética: intercâmbios de energia inter-regionais, geração térmica e metas de níveis de armazenamento dos reservatórios, em função das condições previstas de carga e afluências. As operações diárias são estabelecidas os despachos de geração, intercâmbio e intervenções possíveis e previstas nos equipamentos do sistema para o dia seguinte.

Pode se entender que a Programação da Operação Eletroenergética e a Operação do Sistema possuem grande relação entre elas, visto que a programação da operação para o dia seguinte é um importante dado de entrada para a pré-operação, no qual é realizada num período de tempo imediatamente anterior à operação.

A Operação do Sistema pode ser dividida em 3 subgrupos: a de pré-operação, a de operação em tempo real e a pós-operação, em função de sua dependência com o tempo. A pré-operação realiza consolidar a programação energética diária com a inclusão de modificações no programa de intervenções, nas restrições operativas das instalações de geração e transmissão, na previsão de carga, nas condições previstas para a operação dos reservatórios e na programação do Controle Automático de Geração – CAG. Quanto a operação em tempo real tem como objetivo coordenar, supervisionar e controlar o funcionamento operacional da Rede de Operação, a operação normal do sistema de transmissão e em situações de contingência na rede, e o funcionamento das instalações do SIN. Por fim, as atividades de pós-operação tem como o in-

tuito a apurar os dados de operação realizada, além das análises das ocorrências e perturbações, assim como a divulgação dos resultados para os agentes do setor, os órgãos governamentais, normativos, fiscalizadores e para a sociedade. Além disso, pode se encaixar a normatização da operação, que cuida da formatação de normas e instruções operativas, porém a mesma é realizada sem vínculo temporal definido.

Com isso, pode se destacar que os modelos de predição são importantes para vários segmentos do ONS. Desde o seu planejamento, até a programação da operação eletroenergética e a operação do sistema. Além de se relacionar diretamente com os objetivos do ONS citados, principalmente nos que se referem aos custos de sistema e confiabilidade nos procedimentos de rede para todos. Assim, tendo convicção acerca desses modelos de predição, estes custos relacionados tendem a ser minimizados e a eficiência, por consequência, maximizada.

2.3 MODELOS DE PREVISÃO PARA PRODUÇÃO DE ENERGIA EÓLICA

Diversos modelos já foram construídos para fazer a previsão das séries temporais relacionadas a produção de energia eólica e, segundo Lei et al. (2009), Foley et al. (2012), Holttinen, Miettinen e Sillanpää (2013) e Daraeepour e Echeverri (2014), pode-se separá-las em 4 metodologias:

- (a) métodos físicos (meteorológicos);
- (b) estatísticos tradicionais;
- (c) inteligentes;
- (d) híbridos.

Além disso, também pode-se separar modelos que utilizam uma ou mais de uma variável para a predição. Assim, nas seções 2.3.1, 2.3.2 e 2.3.2.5 serão abordados mais especificamente acerca dos modelos estatísticos tradicionais, inteligentes e híbridos, respectivamente, que foram utilizados no trabalho.

2.3.1 Estatísticos tradicionais- modelos Box-Jenkins (*ARIMA*)

De acordo com Morettin e Tolo (2006), uma série temporal é composta por quatro elementos: 1) Tendência: verifica o sentido de deslocamento da série ao longo de vários tempos; 2) Ciclo: movimento ondulatório que ao longo de vários anos tende a ser periódico; 3) Sazonalidade: movimento ondulatório de curta duração, em geral, inferior a um ano. Está associada, na maioria dos casos, a mudanças climáticas, e; 4) Ruído aleatório ou erro: compreende a variabilidade intrínseca dos dados.

De acordo com Moreira et al. (2010), os modelos de séries temporais podem ser separados em duas formas: modelos de análise da frequência e modelos de análise do tempo. Isto é, modelos nos quais se analisa a quantidade da variação dos pontos em função do tempo, e modelos onde a análise é feita sobre a quantidade dos pontos que é expressa em cada faixa dada de uma série de frequências. Análises de dados de séries temporais servem para obter propriedades e informações de relevância estatística dos dados, enquanto previsões de dados possibilitam a estimação de valores futuros, através de modelos apropriados.

Uma série temporal pode ser entendida como observações de determinada variável, observadas sequencialmente em um determinado período. Tubino (2000) considera que as previsões baseadas em séries temporais são os métodos mais simples e usuais de previsão e que partem do princípio de que um valor no futuro terá um comportamento parecido com seus valores passados. Diferentemente de amostras aleatórias simples, a disposição das observações em função do tempo define uma correlação entre os pontos: teoricamente, quanto mais perto os pontos, maior o grau de dependência entre eles. Esta ordenação do tempo implica que dados que sofreram influência num período passado continuam sob tais influências. Logo, pode-se estabelecer um padrão que determina dados futuros podem de certa forma ser estimados pelos valores passados.

Uma série temporal tem os dados coletados sequencialmente ao longo do tempo e espera-se que ela apresente correlação seriada no tempo. Os modelos de Box-Jenkins, genericamente conhecidos por *ARIMA*- (*Auto Regressive Integrated Moving Averages*) e na literatura em português por Auto Regressivos Integrados de Médias Móveis é um tipo de modelo para ajustes de séries temporais e são modelos matemáticos que visam captar o comportamento da correlação seriada ou autocorrelação entre os valores da série temporal, e com base nesse comportamento realizar previsões futuras. Se essa estrutura de correlação for bem modelada,

fornecerá boas previsões. (WERNER; RIBEIRO, 2003).

Ainda existem na metodologia os modelos ARMA- Auto Regressivos de Médias Móveis (*Auto Regressive Moving Average*) e para a utilização de dados com comportamentos sazonais para o *ARIMA*, utilizam-se o modelo sazonais multiplicativos, o *SARIMA*. Os modelos Garch também são modelos para a previsão de séries temporais. Este é utilizado, em sua maioria, em séries financeiras, nas quais a serie tem distribuição log-normal, com média e variância constantes (CALDEIRA; SOUZA; MACHADO, 2010)

O modelo *ARIMA* (Autoregressivo Integrado e de Média Móvel), desenvolvido por George Box e Gwilym Jenkins, se baseia no ajuste dos valores observados, para assim, reduzir o erro das predições com os valores observados para próximo de zero. Com o modelo *ARIMA*, pode se trabalhar com séries temporais estacionários, ou seja, quando a média e a variância são constantes e o valor da covariância depende da diferença entre os dois períodos do tempo, e também a possibilidade de trabalhar com séries não estacionárias.

Pode se entender que os estes modelos possuem uma extensa variedade de formas; o seu mais usual não sazonal é o *ARIMA* (p,d,q), sendo o p se relaciona com o número de parâmetros autoregressivos, o d o número de diferenciação necessária para deixar a série estacionária e o q o número de parâmetros de médias móveis. Segundo Morettin e Tolo (2006), a construção deste modelo (p,d,q) é dividida em três etapas: identificação, posteriormente a estimação e, por fim, a verificação. Tais etapas se dão a partir da observação da própria série de dados.

Primeiro, verificar qual versão dos modelos de Box-Jenkins descrevem melhor o comportamento da série estudada, após a verificação das funções de autocorrelação (ACF) e das funções de autocorrelação parcial (PACF). A função de autocorrelação mostra o quão forte o valor observado de hoje está correlacionado com os valores do passado. A função de autocorrelação parcial mostra a correlação entre a variável no instante t e uma de suas defasagens, posteriormente, estimar os parâmetros (p,d,q), e, por fim, avaliar a adequação do modelo e se o mesmo descreve o comportamento dos dados de forma correta (WERNER; RIBEIRO, 2003). Assim, a combinação da parte regressiva, mais o grau de diferença e o grau de parte média móvel fica expressa na seguinte equação, que apresentam os valores passados acrescidos dos erros passados:

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \quad (2)$$

no qual Y é a série temporal, e ϕ é o filtro autoregressivo, e_t é um processo puramente aleatório

com média zero e variância constante e θ é o filtro de médias móveis.

Assim, com o Autoregressivo (AR), acrescidos das médias móveis (MA) e a diferenciação, monta-se em conjunto o *ARIMA*.

2.3.2 Modelos inteligentes

Antes de abordar os modelos inteligentes *Random Forest* e *XGBoost*, além dos métodos *ensemble*, apresentar-se-à os modelos *Classification and Regression Trees- CART*, que é a primeira forma de árvore de regressão, nos quais os algoritmos inteligentes foram muito influenciados.

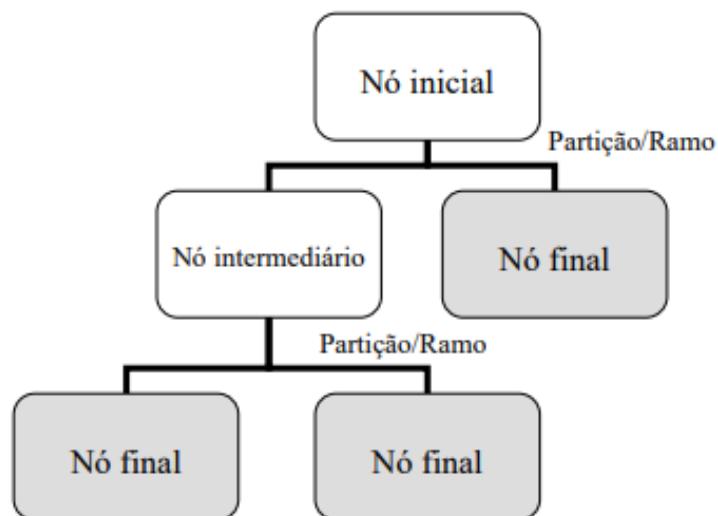
2.3.2.1 *Classification and Regression Trees- CART*

Os modelos *CART* são ótimas alternativas para funções de predição, devido a sua multifuncionalidade e simplicidade de sua árvore. A elaboração destes modelos podem ser utilizadas para a explicação de uma variável resposta numérica, como nos casos de regressão, ou categorizada, nos casos de classificação.

Este método, basicamente, se baseia na execução de sucessivas partições binárias de uma amostra, com base nos resultados das covariáveis, buscando a constituição de subamostras internamente homogêneas.

Existem certos termos que são indispensáveis para o entendimento e caracterização dos componentes de uma árvore de regressão ou classificação, como, nó inicial, nós intermediários e os nós finais (TACONELI, 2008). O nó inicial se relaciona à amostra original, os nós intermediários às subamostras que originam novas subamostras e os nós finais às subamostras não partidas. Também pode ser denotado por t qualquer nó. As partições executadas são denominadas ramos. Por fim, a árvore é a representação gráfica destes nós e ramos, assim como visto na Figura 18.

Figura 18 – Ilustração de uma árvore de classificação/regressão.



Fonte: (TACONELI, 2008).

Segundo Taconeli (2008), a construção destas árvores de regressão podem ser divididas em quatro etapas: definição (e execução) de um critério de partição das amostras, aplicação do processo de poda, seleção do melhor modelo e classificação dos nós finais.

Assim, pode se assumir que a árvore é o resultado final de partições binárias sucessivas de uma amostra original. Estas partições devem ser executadas com base nos resultados para as covariáveis, buscando assim reduzir ao máximo as variações dos resultados nas subamostras constituídas em relação à variação avaliada no nó original. Com isso, devem-se estabelecer critérios específicos para comparação e seleção destas partições amostrais.

Conforme Taconeli (2008), tais partições se dão de maneira diferente em covariáveis ordenáveis ou não ordenáveis. Assim, seja $\{Y_j, X_j\}$, onde $j=1, \dots, n$, uma amostra de tamanho n de uma variável dependente Y e de um vetor p -dimensional de covariáveis $X = (X_{1j}, X_{2j}, \dots, X_{pj})$. Considere X_1 uma variável e T um dos resultados amostrados para X_1 . Com isso pode dividir a amostra em duas distribuindo elementos aos nós distintos conforme a resposta positiva ou negativa à questão " $X_{1j} < T$ ". Repete-se o procedimento para os demais resultados amostrados.

Ainda de acordo com Taconeli (2008), certas restrições devem ser consideradas para a partição do nó. Por exemplo, é indicado que determinar o número mínimo de elementos para que um nó possa ser partido. Subdividir nós com reduzido número de elementos pode aumentar a complexidade do modelo e, assim, afetar na capacidade de predição do mesmo,

devido a instabilidade dos nós formados. Em contrapartida, problemas semelhantes podem ocorrer com nós com poucos elementos. Seguindo, tem-se assim, para as p covariáveis, um conjunto de possíveis partições, que devem ser comparadas, executando-se aquela responsável por uma maior redução da heterogeneidade dos nós. Para tanto, destaca-se algo importante para tais árvores: a medida de heterogeneidade, a fim de quantificar a variação dos dados dentro de um nó t qualquer. Neste sentido, Breiman et al. (1984) destacam diferentes coeficientes dos quais o denominam por medidas de impureza.

O índice ANOVA é uma das sugestões para utilização das árvores de regressão. Este índice é a soma de quadrados dos desvios dos elementos presentes no nó t em torno de sua média.

Como é preferível que se tenha um modelo parcimonioso é indispensável que exista uma árvore de tamanho de nós finais reduzidos. Com isso, se inicia o processo de poda da árvore, conhecido como custo-complexidade (BREIMAN et al., 1984). O processo de poda é terminada com a obtenção da sequencia completa, começada pela árvore original formada até a conjunção dos elementos em um único nó.

Após a escolha da árvore, em suma, nos casos de regressão, a caracterização dos nós finais é feita a partir das médias que contemplam os nós. No método *CART*, utiliza o índice Gini para divisão dos nós. O índice Gini foi proposto por Corrado Gini em 1912 e é comumente utilizado para análises sociais e econômicas.

2.3.2.2 *Extreme Gradient Boosting- XGBoost*

O *XGBoost*, abreviação para *Extreme Gradient Boost*, é uma biblioteca de aprendizado de máquina que pode ser utilizado para problemas de regressão e classificação. Segundo Chen e Guestrin (2016), este método vem ganhando muita popularidade e sendo reconhecido e amplamente utilizado por cientistas de dados, como podem ser observados nas competições da *Kaggle*, no qual grande parte dos participantes vencedores utilizaram este método. O *XGBoost* foi introduzido por Friedman (2001) baseado no método *Gradient Boosted Trees - GBT*. Também pode ser utilizado na resolução de problemas de aprendizagem de máquina supervisionado. O *GBT* é um método pode ser entendido como um método *ensemble* que produz resultados baseado em combinações de árvores de decisão.

Já no *Boosting*, segundo Friedman (2001), que é utilizado no *XGBoost*, tem como objetivo a melhora dos resultados. Assim, o algoritmo baseia-se na ideia de combinar classificadores

genéricos, e menos específicos, para construir classificadores mais robustos.

Além disso, também pode-se verificar a existência da função objetivo, que auxilia o classificador a alcançar o melhor resultado, com auxílio das características e parâmetros. A função objetivo é formada por uma função de perda (*loss function*) somada a uma função de regularização (*regularization function*) (CHEN; GUESTRIN, 2016).

Ao longo das iterações, o *XGBoost* procura encontrar uma árvore de decisão T que minimize a função objetivo. Como pode ser visto na Equação 3.

$$Obj(\theta) = L(\theta) + \Omega(\theta) \quad (3)$$

no qual θ indica os parâmetros do modelo, $L(\theta)$ a função de perda e $\Omega(\theta)$ a função de regularização.

A função de perda objetiva quantificar a precisão do modelo em relação a base de treinamento e é avaliada iterativamente. Já a função de regularização ($\Omega(\theta)$) trata da complexidade do modelo, de modo que limite a divisão da árvore, e tem como objetivo evitar o *overfitting*. Quanto a função de regularização do *XGBoost*, Chen e Guestrin (2016) apresenta a Equação 4.

$$\Omega(\theta) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (4)$$

onde T é a quantidade de folhas na árvore em questão, e representa a soma dos pesos ω_i de cada folha da árvore. Então, o modelo finaliza-se com todas as camadas das árvores de decisão, na qual as folhas dessas árvores tem um valor agregado em cada uma delas. Chen e Guestrin (2016) explica em detalhes os procedimentos do algoritmo em seu artigo.

2.3.2.3 *Random Forest*

O *Random Forest* foi desenvolvido por Leo Breiman e, segundo Díaz-Uriarte e Andres (2006) e Breiman (2001) é um algoritmo de aprendizagem de máquina que utiliza o método *ensemble* em suas camadas finais da regressão em árvore. É um método que pode ser utilizado para regressão e classificação (BREIMAN, 2001). O algoritmo utiliza de uma floresta de árvores formadas através de entradas aleatórias, com ou sem reposição nas mesmas, e, por fim, a combinação dessas árvores por meio da média.

O algoritmo gera, para a k -ésima árvore, um vetor aleatório Θ_k , que independe dos vetores passados $\Theta_1, \dots, \Theta_{k-1}$, podendo ser com reposição, porém, seguem a mesma distribuição.

A árvore é cultivada utilizando este conjunto de treinamento Θ_k , resultando no classificador $h(\tilde{x}, \Theta_k)$, no qual \tilde{x} é um vetor de entrada. Após gerarem um número determinado e suficientemente grande de árvores de decisão, elas são combinadas pela média. Como pode se ver na Equação (5), que demonstra esse processo de decisão, no qual y é a previsão para o conjunto de amostras \tilde{x} e n o total de árvores de decisões geradas.

$$y = h(\tilde{x}) = \frac{1}{n} \sum_{i=1}^n h_i(\tilde{x}) \quad (5)$$

Basicamente, o *Random Forest* gera múltiplas árvores, diferente do modelo CART, que gera apenas uma árvore de decisão. Assim, a "floresta", ou seja, com a soma destas árvores e seus nós, faz-se a média das diferentes saídas.

O algoritmo está disponível em pacotes em vários softwares, inclusive no R, Python, Julia, WEKA, entre outros. Breiman (2001) explica em detalhes os procedimentos do algoritmo em seu artigo.

2.3.2.4 Principal diferença entre o XGBoost e o Random Forest

É consensual na literatura que ambos os métodos são bons algoritmos de regressão em árvores. Sua diferença se dá, basicamente, na forma de aproximação de cada um deles. O *Random Forest* utiliza-se do método *Bootstrap*, enquanto que o modelo *XGBoost* utiliza-se do *AdaBoost*.

2.3.2.5 Ensemble

O método *ensemble* pode ser entendidos como combinações de modelos. Essas combinações podem ser obtidas das mais variadas formas, podendo ser simples, como uma média aritmética, até algo mais robusto, como a junção de características de uma rede neural ou um algoritmo inteligente.

Porém, é importante que esses modelos, além de não correlacionados entre si, possuam características e princípios teóricos de generalização e aprendizado de máquina diferenciados, para assim, formar um *ensemble* mais robusto e parcimonioso. Deste modo, espera-se que o modelo *ensemble* contemple e descreva de forma mais efetiva o banco de dados de interesse, conseguindo resultados melhores do que os modelos individuais usados para a construção do

mesmo.

2.3.3 Seleção de variáveis

O modelo de seleção de entrada funciona a partir da inclusão de todas as variáveis que podem ser importantes para a previsão do vento a partir de seu histórico de informações, além de outros parâmetros que podem ser úteis, como a velocidade, pressão, temperatura, entre outros (DARAEPOUR; ECHEVERRI, 2014). Após a inserção de todas as variáveis possíveis, o modelo apresentará quais serão mais importantes para a previsão desejada.

Assim, pode-se buscar o máximo de informações relevantes para a construção do modelo e descartar as informações que são pouco relevantes, ou irrelevantes, além das informações redundantes, nas quais atrapalhariam na construção do modelo preditor.

A técnica de seleção de variáveis é imensamente usada por diversos algoritmos de *machine-learning* para facilitar a interpretação dos dados, acelerar o processo do modelo, generalizar a performance do mesmo (DARAEPOUR; ECHEVERRI, 2014). Assim como afirma Amjady e Daraeepour (2009), a inclusão de variáveis irrelevantes, além de dificultar e até deixar mais lento o modelo, não propõem nenhuma informação nova importante que melhore a acurácia do mesmo, muitas vezes até prejudicando tal acurácia. Ou seja, de acordo com Kohavi e John (1997), é importante que se utilize um método que funcione como um filtro, retirando os dados que podem atrapalhar na previsão e performance do modelo.

2.3.3.1 Relevância

Basicamente, quando se propõem seleção por relevância, entende-se por selecionar as variáveis individualmente, segundo sua relação com a variável resposta. Ou seja, quanto maior a relevância da variável, maior será seu impacto na variável resposta. Assim, se a relevância for fraca, fraca será a relação entre a variável com a variável resposta. Caso não tenha nenhuma relação, ela será irrelevante, e portanto, pode ser retirada logo na primeira etapa.

Porém, deve se tomar cuidado no momento de escolher a técnica para a escolha dos critérios mais relevantes. De acordo com Daraeepour e Echeverri (2014), diversos métodos podem ser utilizados com este intuito, como *Mutual Information- MI* (LIU; PARK; PRINCIPE, 2009) (AMJADY; DARAEPOUR; KEYNIA, 2010) e medidas de correlação (YU; LIU, 2004).

No trabalho, o *MI* foi escolhido, pois, segundo Amjady e Daraeepour (2009) e Amjady e Keynia (2009), apresenta uma performance melhor na seleção de variáveis pois tem a capacidade de mensurar informações quando a relação entre as variáveis são não lineares.

O *MI* entre duas variáveis x e y pode ser interpretado como a informação que pode ser retirada em y a partir do estudo e observação em x . E o oposto também pode ser percebido. Ou seja, a informação que pode ser obtida em x a partir do que se sabe em y (DARAEPOUR; ECHEVERRI, 2014).

Portanto, denota-se a informação das variáveis x e y como $I(x;y)$. O $I(x;y)$ é definido com base na distribuição conjunta $P(x,y)$ e as respectivas distribuições individuais $P(x)$ e $P(y)$ como segue, no caso discreto:

$$\sum_{i=1}^n \sum_{j=1}^m P(x_i, y_j) \log_2 \frac{P(x,y)}{(P(x)P(y))} \quad (6)$$

O *MI*, apesar de ser um bom método selecionador, não pode ser a única metodologia de seleção. Devido a seleção do mesmo se dar a partir da observação das variáveis individualmente, elas podem ser correlacionadas e serem, assim, redundantes. Ou seja, utilizar as variáveis com maior relevância em um modelo de predição pode não ser a melhor escolha, pois não existiu, ainda, um filtro para selecionar as menos redundantes. Assim sendo, o modelo teria "muito do mesmo", ou seja, praticamente a mesma informação em variáveis diferentes.

2.3.3.2 Importance- Random Forest

O método *VarImportance* do pacote *Random Forest* no software R tem como intuito mostrar quais foram as variáveis que mais contribuíram para a explicação da variável resposta do modelo, visto seu modo de aproximação. A primeira medida a ser calculada pelo método é a partir de permutação de dados *OOB- out-of-bag*: para cada árvore, o erro de previsão na parte de *out-of-bag* dos dados é registrada. Então o cálculo é feito individualmente para cada variável preditora.

A diferença entre a predição e o valor real, em média, é calculada sobre todas as árvores e normalizado pelo desvio padrão das diferenças.

Ele é calculado após a construção das N árvores h_i , e, de acordo com Breiman (1996) e Lahouar e Slama (2015), pode ser calculado da seguinte forma: considerando que o conjunto de treinamento T possui k instâncias (x_j, y_j) , para cada uma dessas instâncias, agrupa-se as árvores h_i que foram construídas com um conjunto T_i . Assim, será obtido k conjuntos de árvores,

denotado por hob_j . Em seguida, é feita uma estimativa \hat{y}_j para cada uma das k instâncias (x_j, y_j) usando seu respectivo conjunto de árvores hob_j .

Assim, o algoritmo Varimportance faz a diminuição das impurezas de cada nó, dividindo os resultados de todas as saídas das árvores.

2.4 REVISÃO BIBLIOGRÁFICA

Na Seção 2.4 serão abordados acerca do estado da arte dos modelos de previsão da produção de energia eólica.

Segundo Monteiro et al. (2009) os modelos Box-Jenkins, *ARIMA*, foram os primeiros a serem construídos para previsão da velocidade do vento e eram utilizados, principalmente, para previsões a curto prazo. Posteriormente, novas metodologias foram utilizadas, como as redes neurais artificiais, *support vector machines - SVM* e os modelos fuzzy. Sfetsos (2000) em seu trabalho comparou diversos modelos para a predição da produção eólica em escala horária. Além da utilização dos modelos *ARIMA* e Redes Neurais, também foi utilizado o "*Adaptive Neuro-Fuzzy Inference Systems* e *Neural Logic Networks*. O modelo das redes neurais, na qual utilizou o algoritmo Levenberg Marquardt para treinamento dos dados, apresentou melhor desempenho quanto ao menor erro de predição. O autor também levou em consideração para suas análises o tempo para as predições.

Quanto a previsão da produção de energia eólica, de acordo com Monteiro et al. (2009), surgem novas possibilidades, como o acréscimo de covariáveis, como a direção do vento, velocidade e etc.

Hong et al. (2016) propõe uma nova abordagem para o curto prazo para previsão de energia eólica, onde a velocidade do vento é prevista e usada para prever a energia eólica através de uma curva de potência obtida a partir de dados históricos. É um modelo com múltiplas variáveis nas quais incluem a velocidade do vento, a produção de energia eólica, o ângulo das turbinas, etc. Os dados utilizados foram fornecidos de uma fazenda eólica no norte da China.

Daraeepour e Echeverri (2014) apresentam um modelo de seleção de variáveis, devido, segundo os próprios, a imensidão de covariáveis que podem influenciar na predição do modelo. Após a seleção os dados são treinados e aplicados numa rede *MLP*. Os resultados são testados em duas fazendas eólicas, e as predições e dados acontecem em escala horária. Outro ponto

importante que pode ser ressaltado é a diferença do *MAPE* com o acréscimo dos dados acerca da temperatura no modelo, confirmando a influência da mesma sobre a produção de energia eólica.

Kusiak, Zheng e Song (2009) propõem 5 modelos para previsão a curto prazo, eles são: SVM, rede neural *MLP*, rede neural RBF, *regression tree* e *random forest*. Em seu trabalho utiliza duas formas de metodologia, abordando mais de uma covariável em seus modelos. Segundo o próprio, o modelo *MLP* foi superior aos demais. Todavia, Fugon, Juban e Kariniotakis (2008) apresentam em sua pesquisa sobre o desempenho de diferentes modelos de mineração de dados para previsão de energia eólica. Duas versões de regressão linear foram examinadas: uma é um modelo de regressão simples usado como referência, e o outro consiste em combinar as variáveis de entrada para criar variáveis extras. Os modelos não lineares utilizados foram as Redes Neurais, SVM, *regression trees* e o *Random Forest*. Os modelos foram testados para 3 fazendas eólicas localizadas na França. Com os resultados foi possível observar que os modelos *Random Forest* obtiveram melhor desempenho.

Quanto os modelos híbridos, Chang et al. (2016) propõe um algoritmo *ARIMA-NN* combinando as vantagens de *ARIMA* e a rede radial de função de base radial (RBFNN). O algoritmo é baseado na aplicação de RBFNN para previsão e suas entradas são as mesmas que *ARIMA*. O modelo combina as vantagens de usar *ARIMA* e a rede neural artificial na previsão de energia eólica. Portanto, o *ARIMA-NN* pode gerar uma função não-linear para criar um método preciso para prever séries temporais, para curto-prazo, de velocidade do vento e potência de energia eólica. Segundo Chang et al. (2016) o modelo proposto é superior ao *ARIMA* e a rede neural utilizadas no que se refere ao menor erro de predição, mostrando que a combinação dos modelos pode ser de grande valia para previsão de séries temporais. Para as variáveis de entrada, o autor utiliza a inclinação do vento, velocidade e temperatura.

Jursa et al. (2007) constrói modelos inteligentes para predição de energia eólica um passo a frente. A autora compara diferentes técnicas de previsão, com um MLP NN, SVM entre outros. Além de utilizar um algoritmo para seleção de variáveis. Os resultados de 10 fazendas eólicas na Alemanha são produzidos e comparados. O melhor modelo foi a combinação de três modelos: SVM, *nearest neighbor* e *mixture of experts*. O novo modelo teve um desempenho superior as redes neurais de, aproximadamente, 15%. A combinação de todos os modelos, porém, não foi tão significativa, assim, percebe-se que nem sempre acrescentar mais modelos é a melhor alternativa. Quanto aos modelos individuais, as redes neurais apresentou, na maioria das

fazendas, o pior desempenho. Já o modelo *mixture of experts* apresentou melhor desempenho, seguido do SVM.

Yang et al. (2015) apresentam Modelos de Markov aprimorado pelo *SVM - Support Vector Machine* para a previsão de curto prazo, e os resultados foram extremamente significativos. O autor utiliza os dados históricos das energias das turbinas eólicas em um parque eólico, e aplica os modelos de Markov em um estado finito para ter uma noção das flutuações 'normais' da geração do vento, tendo em conta que a série não é estacionária e possui sazonalidade, posteriormente, o *SVM* é integrado ao modelo de Markov. Os resultados são comparados com os resultados dos modelos de Markov e *SVM*, além de alguns modelos autoregressivos.

Também baseado na combinação de métodos, Li, Wang e Goel (2015) montam vários modelos com base nas redes neurais para depois combinarem, além de utilizar uma técnica de seleção de variáveis, a *Mutual Information*. Além disso, também é utilizado a transformação de 'Wavelet' para trabalhar com os dados por não serem estacionários. Por fim, o novo modelo é comparado com métodos usuais, e se sai melhor nas métricas de avaliação para curto-prazo. São analisados o *MAPE* e o *NRMSE* para 1, 24 e 48 horas, e o modelo é superior aos demais em todas estas comparações.

Pelo que pode ser percebido, os modelos híbridos, ou combinações, são, na maioria dos casos, superiores aos modelos individuais no que se refere ao menor erro de predição. Existem diversas formas nas quais os modelos podem ser combinados, a média é a mais simples.

Como foi visto, a série da produção de energia eólica possui sazonalidade e não é estacionária, o que dificulta na predição dos mesmos. Porém, pode-se relacionar esta produção com a temperatura, assim como afirma Thornton et al. (2017), que existe uma tendência de baixa produção de energia eólica nos momentos de inverno mais denso (temperatura mais baixas), o que pode ser um pouco preocupante, visto que, ao contrário da produção de energia eólica, a demanda elétrica aumenta nestas situações. Com isso, na construção dos dados para as previsões, tão importante quanto os modelos, se referem ao tratamento destes dados. (FUGON; JUBAN; KARINIOTAKIS, 2008)

3 MATERIAL E MÉTODOS

Na Seção 3.1 será abordado acerca do material utilizado para a dissertação. Na Seção 3.2 serão dissertados acerca dos dados históricos da produção de energia eólica. A Seção 3.3 abordará sobre a metodologia proposta.

3.1 MATERIAL

O software R foi utilizado para as análises e modelagens, por ser um software estatístico muito eficiente e gratuito. Quanto aos modelos Box-Jenkins utilizou-se o pacote 'forecast' e 'tseries'. Os modelos de *Random Forest* pelo pacote 'randomForest' e o *XGboost* pelo pacote 'XGboost'.

3.2 DADOS

Os dados contemplam a produção de energia eólica de 2011 a 2016, anotados a cada 5 minutos. Ao longo dos anos teve aumento na produção de energia eólica, porém, não foi um crescimento notável no período de análise, conforme visto na Tabela 3. Os dados foram disponibilizados a partir do <https://transmission.bpa.gov/business/operations/wind/>, e contemplam a geração de energia eólica observada nos Estados Unidos.

Para aplicação dos dados nos modelos, os mesmos foram convertidos para horário.

Tabela 3 – Produção de energia eólica em MW do primeiro semestre de 2011 e 2016.

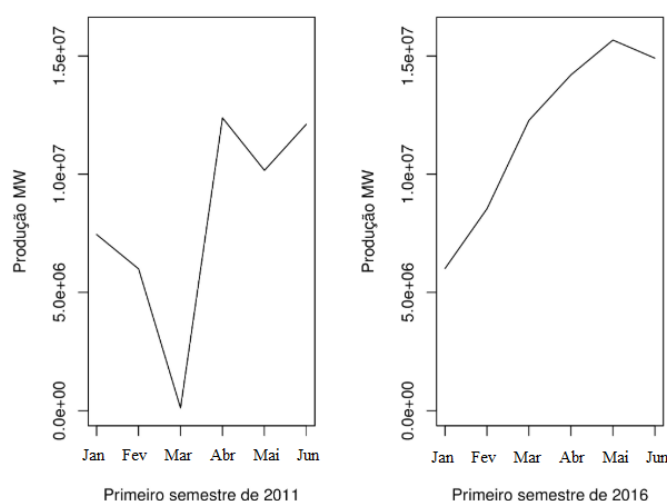
Ano	Máximo em 5 minutos	Média em 5 minutos	Desvio Padrão	Produção total em MW
2011	3339	1101,50	950,32	57409051
2012	4295	1321,57	1189,01	69255741
2013	4512	1381,68	1361,63	72007637
2014	4319	1410,30	1314,58	73499190
2015	4568	1079,43	1245,92	56258355
2016	4564	1366,50	1370,68	71610083

Fonte: Elaborada pelo autor.

Observando o desenvolvimento da produção total ao longo do primeiro semestre de 2011

e 2016, percebe-se algumas características semelhantes, conforme visto na Figura 19. Ou seja, é visto que os primeiros meses a produção de energia eólica é baixa, e nos três meses finais ela é mais alta.

Figura 19 – Produção total do primeiro semestre de 2011 e 2016.



Fonte: Autor.

Quanto aos dados do segundo semestre, como pode ser percebido a partir da Tabela 4 o ano de 2016 também foi superior na produção comparado ao ano de 2011. Também é possível observar que a produção total no segundo semestre dos anos é menor que no primeiro semestre.

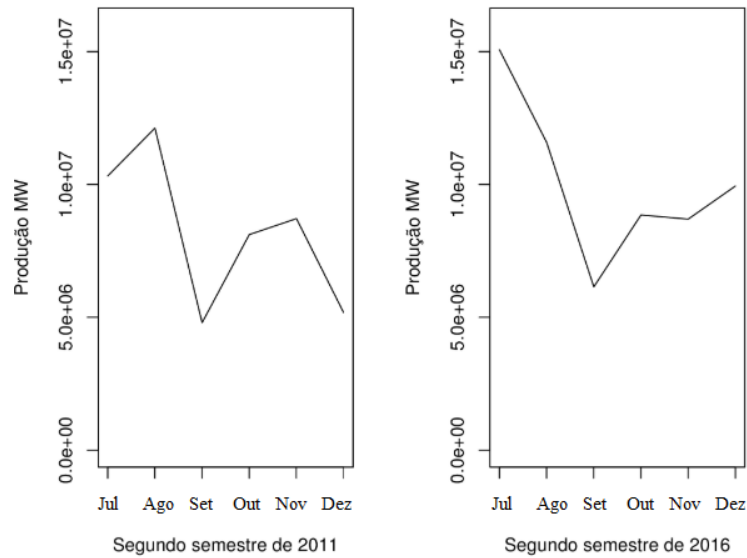
Tabela 4 – Produção de energia eólica em MW do segundo semestre de 2011 a 2016.

Ano	Máximo em 5 minutos	Média em 5 minutos	Desvio Padrão	Produção total MW
2011	3437	958,43	974,07	50800913
2012	4365	1030,57	1132,33	54624445
2013	4290	1141,39	117,91	60498326
2014	4339	1133,77	1232,20	60094643
2015	4594	1313,79	1319,57	69636193
2016	4588	1226,45	1173,04	65007061

Fonte: Elaborada pelo autor.

Assim como no primeiro semestre, as produções do segundo semestre também seguem caminhos semelhantes. Como visto na Figura 20, o ponto mais baixo de produção também se dá no mês de setembro, após uma queda nos meses anteriores.

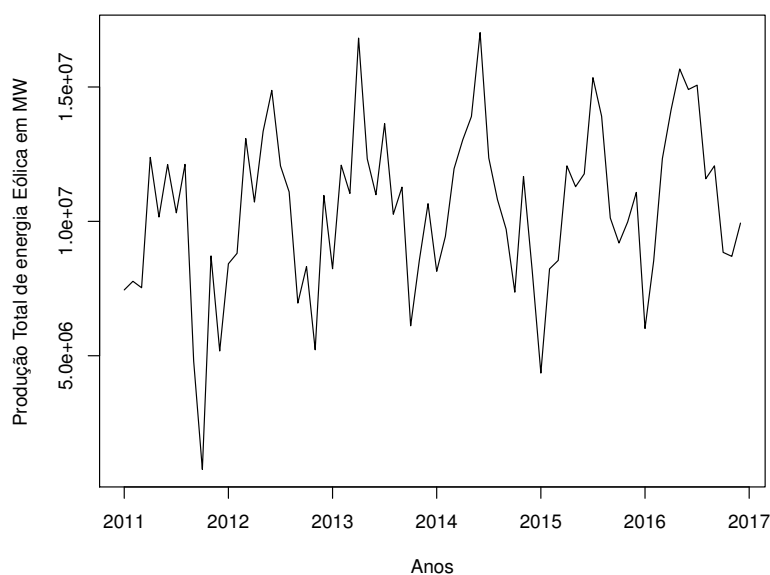
Figura 20 – Produção total do segundo semestre de 2011 e 2016.



Fonte: Autor.

A Figura 21 apresenta o gráfico da série de produção total dos anos de 2011 a 2016, completa. Também pode-se perceber um padrão nos mesmos, percebendo os meses nos quais a produção é maior e menor, apresentando sazonalidade e tendência. Ou seja, percebe-se no comportamento da série que a produção de energia eólica repete o comportamento em certos períodos idênticos do tempo. O que pode destacar a relação da produção de energia eólica com a temperatura e as estações.

Figura 21 – Produção total em MW de 2011 a 2017.



Fonte: Autor.

De acordo com Daraeepour e Echeverri (2014), a análise e construção do modelo pode ser melhor estruturado quando separado por estação, visto a grande relação da produção de energia eólica e velocidade do vento com a temperatura. Assim como pode ser percebido nas Figuras 19, 20 e 21, os meses do ano referentes ao inverno, estação mais fria, a produção eólica é menor.

3.3 METODOLOGIA PROPOSTA

A metodologia proposta será apresentada em quatro tópicos, explicando passo a passo de como foram feitas as predições, assim como a construção dos modelos e a forma de como sucedeu o tratamento dos dados para as construções dos mesmos, além da construção do *ensemble* proposto.

Para verificar os resultados em cada estação do ano, cada modelo citado nas Seções 3.3.1, 3.3.2 e 3.3.3 apresentarão as predições de uma semana, em escala horária, para as quatro estações do ano de 2016, utilizando 1 passo a frente, ou seja, após a predição da hora t , o modelo usará o t real para predizer $t + 1$, e não o t predito. O método de *ensemble* proposto no trabalho

será detalhado para cada estação do ano.

3.3.1 *Random Forest e XGBoost*

Para as predições com os modelos *Random Forest* e *XGBoost* foi utilizado o conjunto de dados de 2014 até 2016, em escala horária para o treino. Optou pela redução do conjunto de dados devido ao tempo de processamento para os modelos inteligentes.

O banco de dados para a predição sem a seleção foi construído com as 336 horas passadas para explicar a variabilidade da hora 337. Ou seja, para predizer a hora t , utilizou-se as horas passadas $t - 336, \dots, t - 1$ como variáveis explicativas. Após a seleção de variáveis, utilizou-se até a quinta hora passada, ou seja, até $t - 5$ para predizer t .

Nos resultados, serão comparados os resultados dos modelos com e sem seleção, para posteriormente compará-los com os outros modelos individuais, juntamente com o *ensemble*.

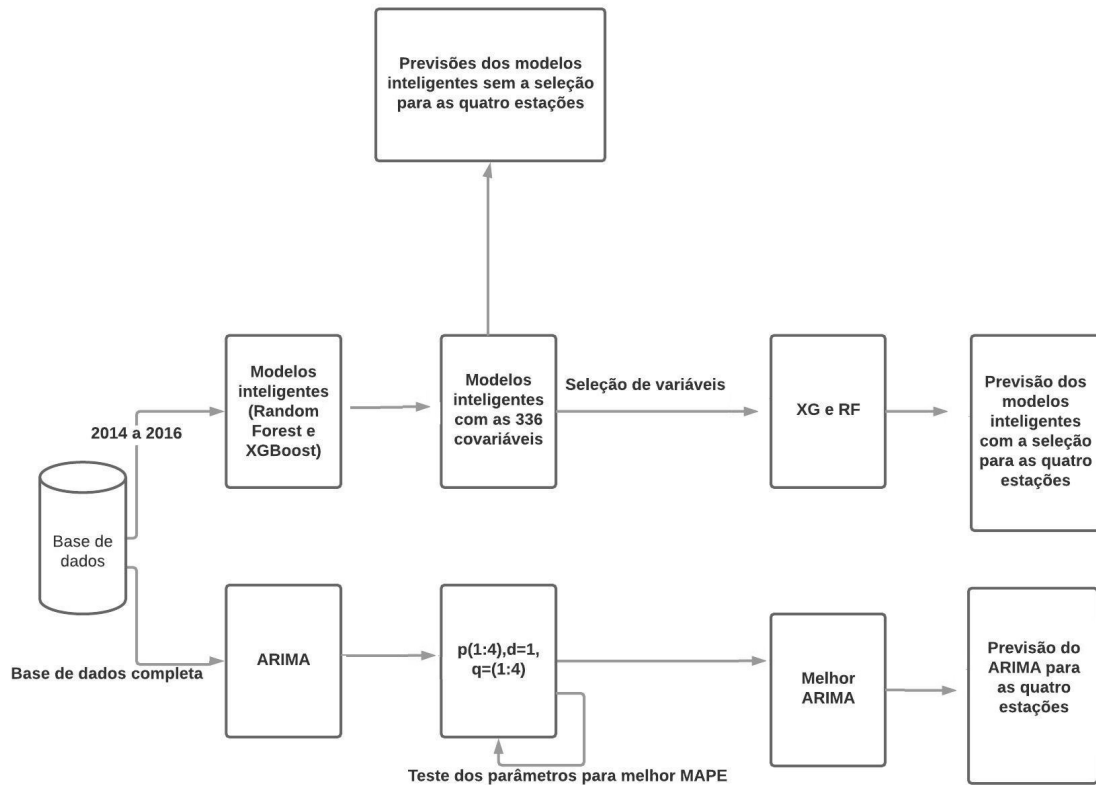
3.3.2 *ARIMA*

Para as predições com o modelo *ARIMA*, utilizou-se o conjunto inteiro dos dados. De 2011 a 2016. Diferente dos modelos individuais, os modelos Box-Jenkins não necessitam da seleção utilizada no trabalho.

Assim, foram construídos diversos modelos com os parâmetros p, d, q . O parâmetro p variou de 1 a 4. O parâmetro q também. Porém, como foi aplicado apenas uma diferenciação, o d foi igual a 1.

Basicamente, a construção dos modelos individuais pode ser visto na Figura 22, no qual aponta, primeiramente a separação em dois do banco de dados para desenvolver os diferentes métodos. No caso dos modelos individuais, primeiro a transformação do banco de dados, seguido da predição da semana alvo sem a seleção de variáveis, e por fim a seleção e sua previsão. No caso do *ARIMA*, primeiro os testes para escolha dos melhores parâmetros, para posteriormente fazer as predições das estações.

Figura 22 – Fluxograma da construção dos modelos individuais.



Fonte: Autor.

3.3.3 Ensemble XG

Basicamente, pode-se dividir a construção do *Ensemble XG* em duas etapas. Na primeira, realiza-se as previsões anteriores com cada modelo individualmente, para por fim, na segunda, acrescenta-las como variáveis explicativas, além da última hora anotada. Assim, espera-se que o novo modelo consiga melhorar os resultados quanto aos erros comuns dos modelos individuais.

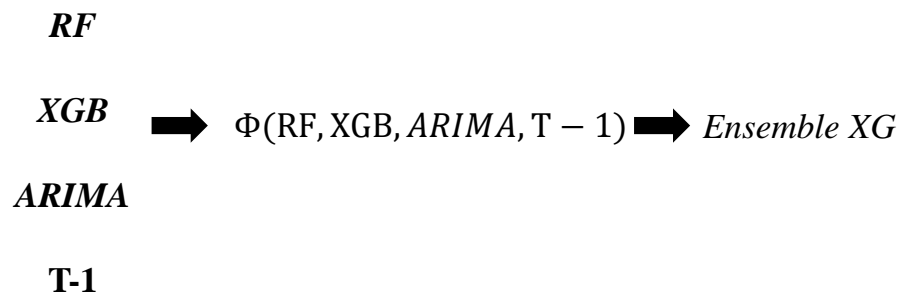
Para a prever cada estação, os dados de treino dos modelos *ensemble* foram diferentes. Basicamente, como pode ver na Figura 23, o *ensemble* é composto pelas saídas dos modelos *Random Forest*, *XGBoost*, *ARIMA*, além da última observação do histórico. Ou seja, a partir das previsões dos modelos individuais e do histórico, o algoritmo *XGBoost* foi utilizado para combiná-los em um novo modelo.

Mais profundamente, decidiu-se optar por 700 horas de previsão de cada modelo, aproximadamente 1 mês, para realizar a combinação, no qual será chamado de *Ensemble XG*. Ou

seja, realizou-se a predição destas 700 horas anteriores da semana alvo com os 3 modelos. Com isso, entende-se que, com estas predições, acrescido do histórico da produção eólica, exista um bom número de período de treino para o *Ensemble XG*, e com isso o modelo entenda o peso ideal para cada variável explicativa e apresente uma predição com menor erro.

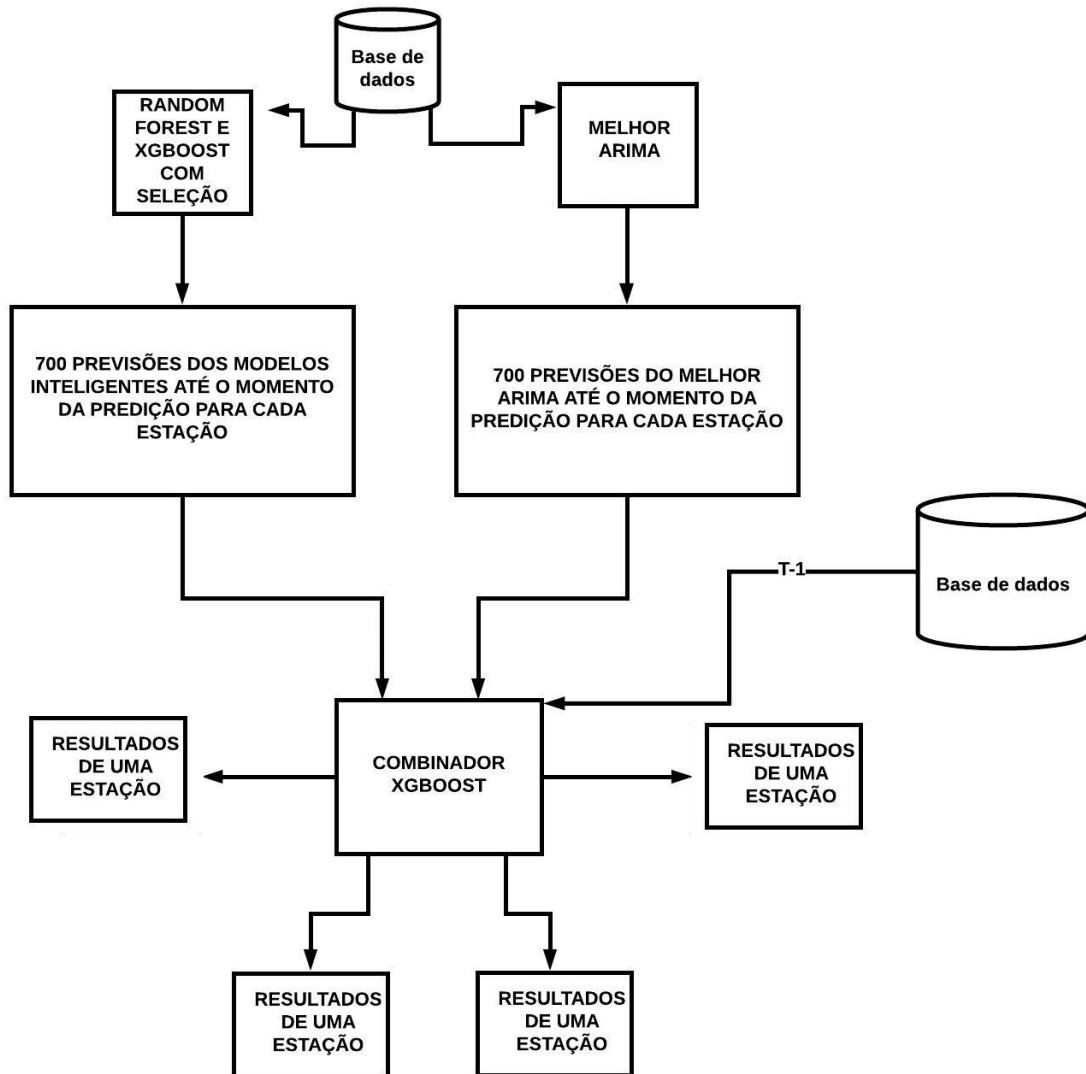
Na Figura 24, apresenta o fluxograma das etapas do *Ensemble XG*. Assim como na Figura 22, a construção do *ensemble* começa com a separação dos dados para o *ARIMA* e para os modelos inteligentes. Após esta etapa, as predições são realizadas para compor a fase de treino e, por fim, combinados para a predição do novo modelo.

Figura 23 – *Ensemble* proposto.



Fonte: Autor.

Figura 24 – Fluxograma do *Ensemble* proposto.



Fonte: Autor.

Quanto ao período de treino, se a semana a ser predita pelo novo modelo começa na hora t , as variáveis explicativas do modelo serão compostas pelas previsões dos modelos individuais da hora $t - 700$ até t , além da observação real da produção de energia eólica na hora $t - 701$ até $t - 1$:

$$\text{TREINO} = \begin{pmatrix} XG_1 & RF_1 & ARIMA_1 & T_0 \\ XG_2 & RF_2 & ARIMA_2 & T_1 \\ \dots & \dots & \dots & \dots \\ XG_n & RF_n & ARIMA_n & T_{n-1} \end{pmatrix}$$

Onde XG_n representa as previsões do *XGBoost* no período n , RF_n as previsões do *Random Forest* no período n , $ARIMA_n$ as previsões do *ARIMA* no período n , e por fim, T_{n-1} as observações reais no período $n-1$.

3.4 MÉTRICAS DE AVALIAÇÃO

As métricas utilizadas para avaliar as performances dos modelos em relação ao objetivo foram a *Mean Absolut Percentage Error* (MAPE) (Equação. 7) e o *Root mean squared error* (RMSE) (Equação. 8).

$$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

onde \hat{y}_i representa a carga prevista e y_i referencia a carga observada.

Ambas as métricas de avaliação vão de 0 a infinito, e quanto menor forem os resultados, melhores são consideradas as previsões quanto ao erro.

4 RESULTADOS E DISCUSSÃO

Na Seção 4.1 será feita uma breve análise exploratória dos dados. Posteriormente, na Seção 4.2 discorrerá acerca do desenvolvimento dos modelos, seus parâmetros, e as variáveis selecionadas, para por fim apresentar os resultados parciais para cada estação e a performance de cada modelo, antes e depois da seleção de variáveis, além do *ensemble*.

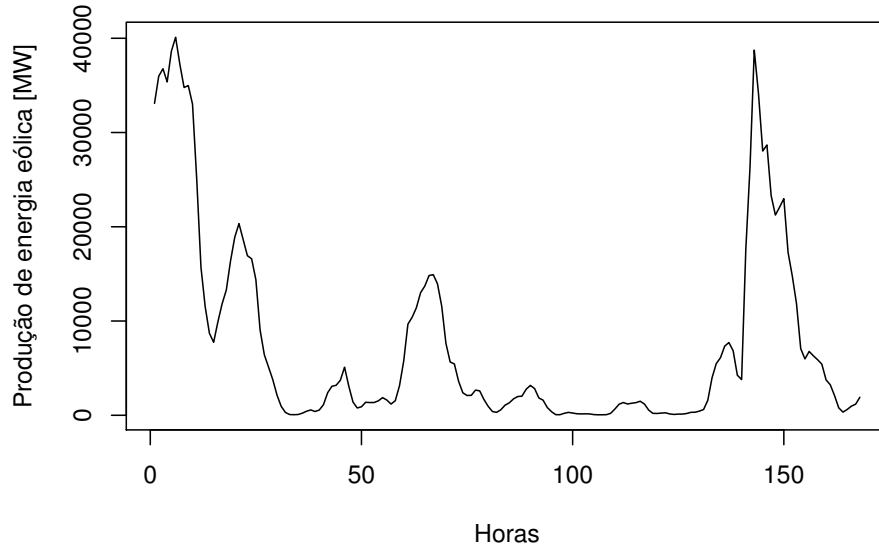
4.1 ANÁLISE EXPLORATÓRIA

As previsões se deram em 4 etapas no ano de 2016: uma semana de previsão para cada estação. Estas semanas foram escolhidas no meio da estação, para não ter influência da estação passada. As análises dos resultados por estação podem ser divididos em duas etapas: na primeira a análise do *MAPE* diário dos modelos individuais inteligentes (*Random Forest* e *XGBoost*). No segundo a análise do *MAPE*, do *RMSE* horário total, além da comparação dos extremos com o banco de dados histórico. As semanas preditas foram:

- (a) Primeira semana alvo: 9 de abril até o dia 15 de abril de 2016;
- (b) Segunda semana alvo: 9 de julho até o dia 15 de julho de 2016;
- (c) Terceira semana alvo: 8 de outubro até o dia 15 de outubro de 2016;
- (d) Quarta semana alvo: 24 de dezembro até o dia 31 de dezembro de 2016.

Na Figura 25, que representa as 168 horas desta primeira semana alvo, percebe-se um comportamento, apesar de alguns picos, com uma produção de energia eólica mantendo-se baixa em boa parte do tempo, em comparação com as demais semanas preditas. Estas diferenças são explicadas devido a sazonalidade e influência de diversas covariáveis na produção deste tipo de energia.

Figura 25 – Os dados referentes a primeira semana alvo para a predição.

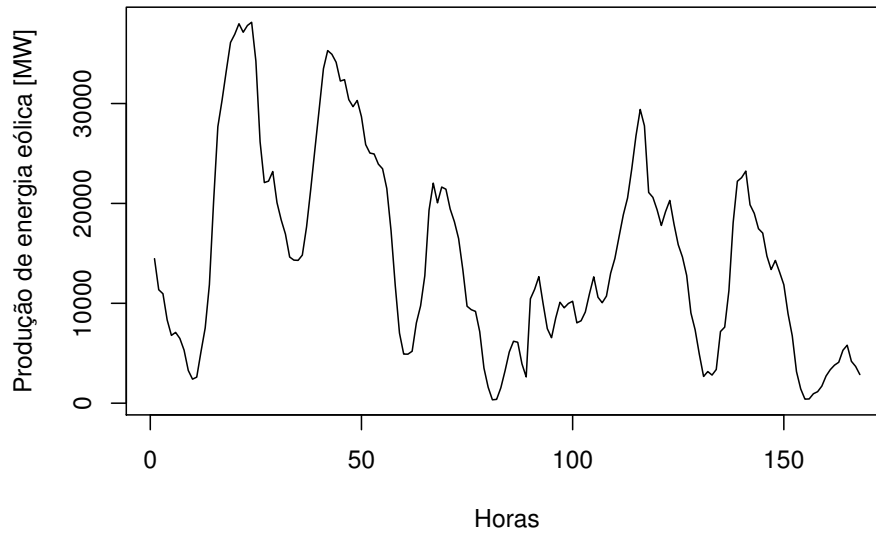


Fonte: Autor.

Os dados referentes a segunda estação a serem preditas contemplam a produção de energia eólica em base horária do dia 9 de julho até o dia 15 de julho de 2016.

O período observado, visto na Figura 26, apresenta uma produção com uma variabilidade maior. Porém, pode-se destacar que, diferente da Figura 25, ela não se mantém em baixa, além dos extremos serem menos intensos.

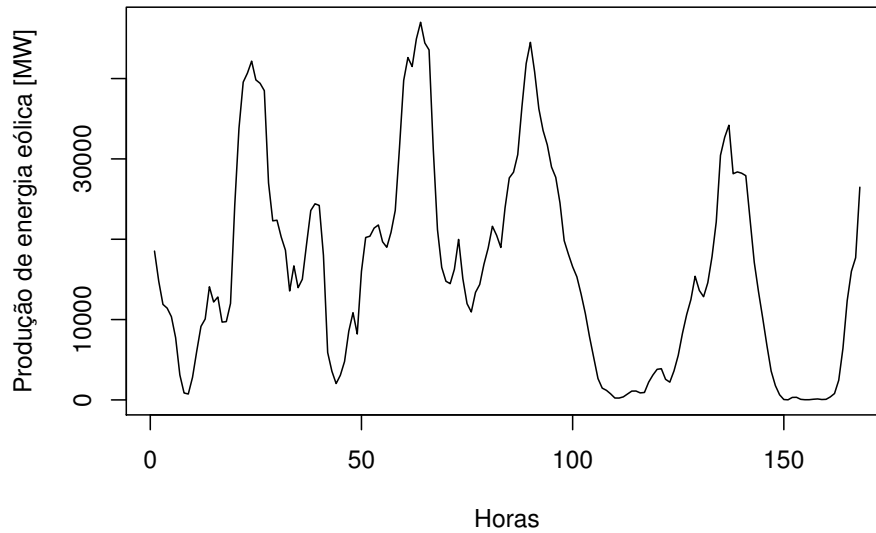
Figura 26 – Os dados referentes a segunda semana alvo para a predição.



Fonte: Autor.

A Figura 27 representa as 168 horas a serem preditos na segunda semana de outubro, ou a terceira semana alvo da predição. Mais especificamente, são referentes ao do dia 8 de outubro até o dia 15 de outubro de 2016. Neste caso, percebe-se que é a semana com maior variabilidade e mais picos do que os demais, o que pode prejudicar os modelos em relação ao *MAPE* e *RMSE*.

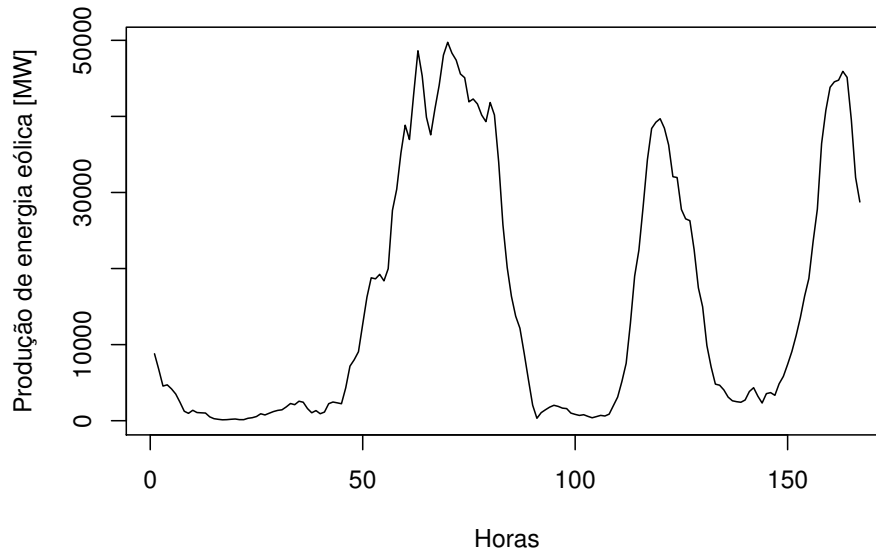
Figura 27 – Os dados referentes a terceira semana alvo para a predição.



Fonte: Autor.

Quanto a Figura 28, apresenta as 168 horas a serem preditos pelos modelos do dia 24 de dezembro até o dia 31 de dezembro de 2016, ou a quarta semana alvo. Assim como na Figura 25, percebe-se que a produção se mantém em baixa em muitas horas, porém, sua diferença se dá nos picos, visto a maior ocorrência dos mesmos, além de seus valores máximos serem maiores.

Figura 28 – Os dados referentes a quarta semana alvo para a predição.



Fonte: Autor.

Na Tabela 5, é visto que a primeira e segunda estação alvo, além da média ser menor, o desvio padrão também é, assim, os modelos tendem a predizer de forma mais precisa tais dados. Quanto as terceira e quarta semana, visto suas maiores variabilidades, os modelos tendem a errar mais, principalmente no que se refere aos extremos, o que pode prejudicar, principalmente, o valor do *RMSE*.

Tabela 5 – Análise descritiva das semanas a serem preditas.

Estações	Média de produção	Produção Total	Desvio Padrão
Primeira estação	7478,40	1256371	10327,37
Segunda estação	14700,07	2469611	10011,38
Terceira estação	16217,40	2740741	12739,67
Quarta estação	15534,14	2625270	16525,10

Fonte: Elaborada pelo autor.

Analisando os dados, comprova-se a existência dos diferentes comportamentos ao longo dessas semanas alvo nas estações, reforçando o que se sabia acerca das influências das estações. Também pode-se relacionar as mudanças de baixa e alta produção de energia eólica com as estações e períodos mais frios do ano.

4.2 RESULTADOS DOS MODELOS

Para facilitar na construção do modelo e diminuir a variabilidade e ruído branco da série optou-se por transforma-la em dados horários, além de ser mais usual o tratamento dos dados nesta escala, tanto para trabalhos acadêmicos, como para as previsões em distribuidoras de energia.

Cada modelo construído teve sua especificidade, assim, ao longo dos resultados, destacar-se-á cada um deles. Porém, todos os modelos fizeram a previsão um passo a frente. Ou seja, após cada previsão, os modelos foram realimentados com os dados reais observados, assim, para prever t , utilizou-se o histórico até $t - 1$. Assim, para prever o $t + 1$, em vez de utilizar o valor predito de t , utilizou-se o valor real, e assim por diante. Os resultados foram separados para cada estação.

Para a seleção de variáveis se optou por dois métodos: *Mutual Information* e o *VarImportance* do pacote *randomForest*.

4.2.1 Parâmetros para a construção do modelo *Random Forest*

Para a construção do modelo *Random Forest* utilizou-se 5 parâmetros: *ntree*, *nodesize*, *mtry*, *nPerm* e *corr.bias*. Os mesmos foram escolhidos a partir do melhor desempenho na predição de períodos no ano de 2015.

Tabela 6 – Parâmetros para as construções dos modelos *Random Forest*

<i>Random Forest</i>				
<i>ntree</i>	<i>nodesize</i>	<i>mtry</i>	<i>nPerm</i>	<i>corr.bias</i>
500	12	9	1	6

Fonte: Elaborada pelo autor.

- (a) ***Ntree***: Número de árvores para crescer. Isso não deve ser definido para um número muito pequeno, para garantir que cada linha de entrada é prevista pelo menos algumas vezes. Porém, um número muito alto pode causar *overfitting*.
- (b) ***Nodesize***: Tamanho mínimo dos nós terminais. Definir este número maior faz com que árvores menores sejam cultivadas (e, portanto, levem menos tempo).

- (c) **Mtry**: Número de variáveis amostradas aleatoriamente como candidatos em cada divisão.
- (d) **NPerm**: Número de vezes que os dados *out of bag*- OOB são permutados por árvore para avaliar a importância da variável. Número maior que 1 fornece uma estimativa ligeiramente mais estável, mas não muito eficaz. Mais comumente usado para regressão.
- (e) **Corr.bias**: Correção de viés para regressão.

Quanto ao tempo de execução de cada modelo, os *Random Forest* o desempenho do modelo pode ser visto na Tabela 7:

Tabela 7 – Tempo de execução médio dos modelos *Random Forest* em segundos para a predição de uma hora.

Modelos	Usuário	Sistema	Decorrido
<i>Random Forest</i>	180,98	0,04	181,09
Random Forest sem seleção	503,08	0,19	503,83

Fonte: Elaborada pelo autor.

Segundo a função *system.time()* o tempo total decorrido para a predição de uma hora nos modelos *Random Forest* com e sem a seleção foram de 181,09 e 503,83 segundos, respectivamente. Assim, percebe-se uma diminuição significativa no tempo de processamento dos mesmos com a diminuição das variáveis.

4.2.2 Parâmetros para a construção do modelo *XGBoost*

Para a construção dos modelos *XGBoost* foram utilizados 7 parâmetros e 1 especificação: *nrounds*, *nthread*, *max_depth*, *eta*, *min_child_weight*, *subsample* e *colsample_bytree*, além de *objective* como especificação. Assim como no caso do *Random Forest*, os parâmetros foram escolhidos a partir dos testes de previsões no ano anterior, 2015.

Tabela 8 – Parâmetros para as construções dos modelos *XGBoost* 1.

<i>XGBoost</i>				
<i>nrounds</i>	<i>nthread</i>	<i>max_depth</i>	<i>eta</i>	<i>min_child_weight</i>
500	2	10	0,02	9

Fonte: Elaborada pelo autor.

Tabela 9 – Parâmetros para as construções dos modelos *XGBoost* 2.

<i>XGBoost</i>		
<i>subsample</i>	<i>colsample_bytree</i>	<i>objective</i>
0,2	1	"reg:linear"

Fonte: Elaborada pelo autor.

- (a) ***Nthread***: número de encadeamentos paralelos usados para executar o *XGBoost*
- (b) ***Eta*** : encolhimento do tamanho do passo usado na atualização para evitar *overfitting*. O *eta* reduz os pesos dos recursos para tornar o processo de reforço mais conservador. Torna entre 0 e 1.
- (c) ***Max_depth***: profundidade máxima de uma árvore de regressão, aumentar esse valor tornará o modelo mais complexo, porém pode causar um *overfitting*.
- (d) ***Min_child_weight***: soma mínima do peso da instância (hessian) necessária em uma *child*. Se a etapa de partição da árvore resultar em um nó folha com a soma do peso da instância menor que *min_child_weight*, o processo de construção dará mais particionamento. No modo de regressão linear, isso simplesmente corresponde ao número mínimo de instâncias necessárias em cada nó. Quanto maior, mais conservador será o algoritmo.
- (e) ***Subsample***: proporção da *subsample* da instância de treinamento. Utilizar com um *subsample* < 0,5 pode evitar o *overfitting*.
- (f) ***Colsample_bytree***: proporção de *subsample* de colunas para construção de cada árvore.
- (g) ***Nrounds***: número máximo de interações.

Tabela 10 – Tempo de execução médio dos modelos *XGBoost* em segundos para a predição de uma hora.

Modelos	Usuário	Sistema	Decorrido
<i>XGBoost</i>	17,47	2,92	12,94
XGBoost sem seleção	869,92	7,63	454,53

Fonte: Elaborada pelo autor.

Segundo a função *system.time()* o tempo total decorrido para a predição de uma hora nos modelos *XGBoost* com e sem a seleção foram de 12,94 e 454,53 segundos, respectivamente. Pode-se observar que os modelos *XGBoost* foram mais rápidos do que os modelos *Random Forest*, comparando os dados com e sem a seleção. Porém, vale destacar que o tempo

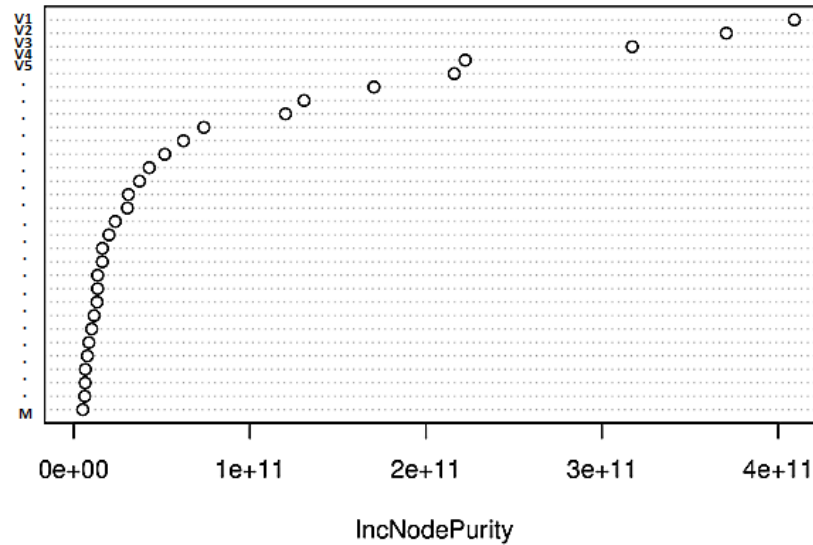
decorrido do modelo *XGBoost* com a seleção aumentou 3504% enquanto que esse aumento no caso do *Random Forest* foi de 278%. Assim, consta-se que, para menos variáveis, o *XGBoost* tem o seu processamento mais acelerado, enquanto que, com o acréscimo de mais variáveis, o *Random Forest* tende a aumentar menos o seu tempo de processamento (percentualmente) em comparação ao *XGBoost*.

4.2.3 Variáveis selecionadas

A princípio, utilizou-se todas as 336 covariáveis para explicar o modelo, ou seja, as 336 horas passadas foram adicionadas como covariáveis para a predição da produção de energia eólica. Após a seleção que se deu a partir das mais relevantes pela *Mutual information* e a *VarImportance*, do pacote *randomForest*, o modelo passou a contar com apenas 5 covariáveis: as 5 horas passadas. Assim, segundo estes métodos de seleção, as 5 últimas horas são as mais importantes para prever a produção eólica.

Conforme mostra a Figura 29, cujo o eixo y são as variáveis do modelo e o eixo x a importância segundo o *VarImportance*, pode se observar que a primeira observação passada é a mais importante para a predição. Além disso, também se percebe uma diferença de importância a partir da terceira observação passada em relação a quarta e quinta observação.

Figura 29 – VarImportance dos dados totais com as 336 covariáveis.

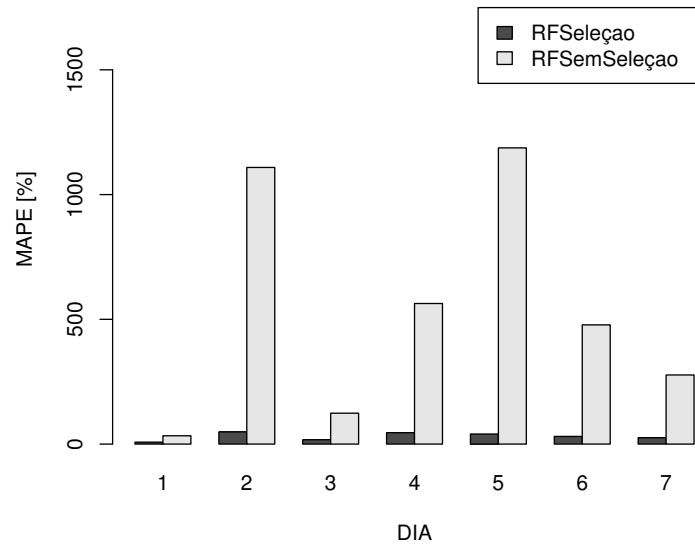


Fonte: Autor.

Quanto o *Mutual Information*, também pode se perceber uma discrepância nas primeiras observações, porém, percebe-se que até a quinta observação tem uma relevância maior. Visto isso, optou-se pelas 5 primeiras horas passadas.

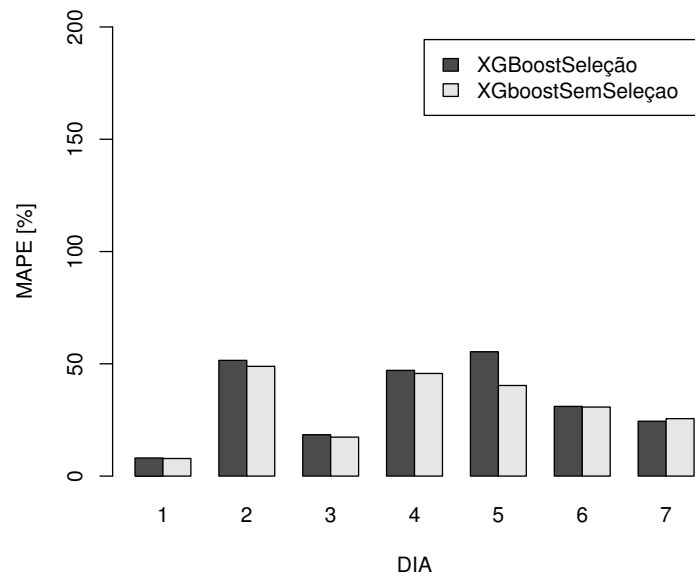
4.2.4 Resultados para primeira estação

Como percebe-se na Figura 30, a seleção de variáveis, no qual se obteve pelas 5 últimas horas observadas como covariáveis, aplicadas para prever a primeira semana alvo foi extremamente efetiva. Apenas no primeiro dia de previsão que o modelo do *Random Forest* não foi significativamente pior, porém, nos outros dias, a diferença do *MAPE* foi muito relevante.

Figura 30 – MAPE dos dias do *Random Forest* na primeira estação.

Fonte: Autor.

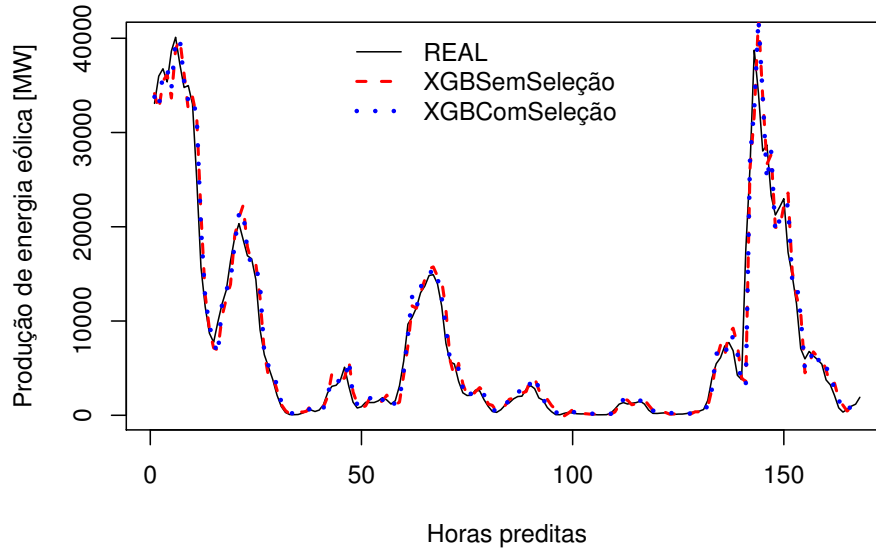
Quanto ao modelo *XGBoost*, é percebido que a seleção de variáveis não obteve o mesmo impacto na acurácia dos resultados como no *Random Forest*, porém, vale ressaltar que, em média, o *MAPE*, conforme visto na Figura 31 e na Tabela 11, diminuiu, o que constata a eficácia de tal seleção.

Figura 31 – MAPE dos dias do *XGBoost* para a primeira estação.

Fonte: Autor.

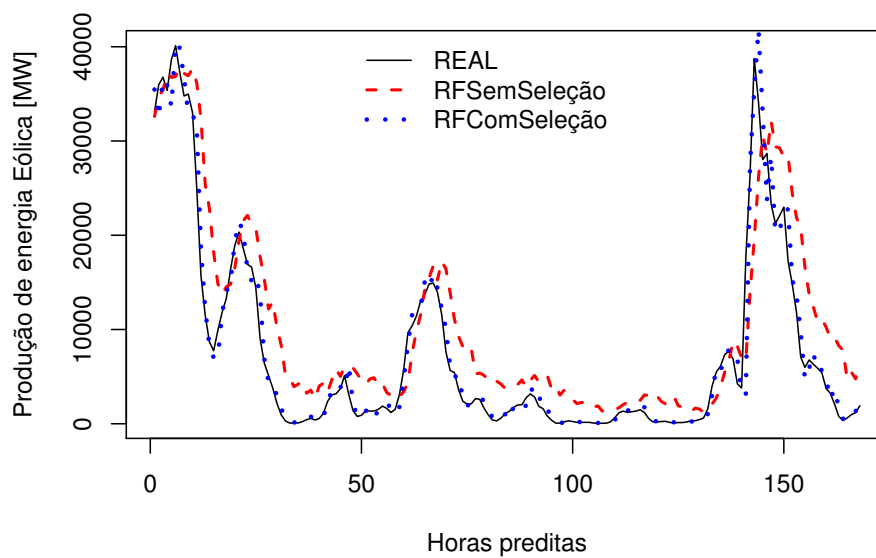
Tanto o modelo *Random Forest*, como o *XGBoost*, mesmo após a seleção de variáveis, na predição dos extremos sempre supervalorizaram, principalmente no último pico, no qual é a maior produção de energia eólica observada, e também, o maior erro dos modelos, como pode ser visto nas Figura 32 e 33. Vale destacar que o modelo *Random Forest* sem a seleção teve um desempenho bem abaixo dos demais.

Figura 32 – Predição dos modelos *XGBoost* na primeira estação.



Fonte: Autor.

Figura 33 – Predição dos modelos *Random Forest* primeira estação.

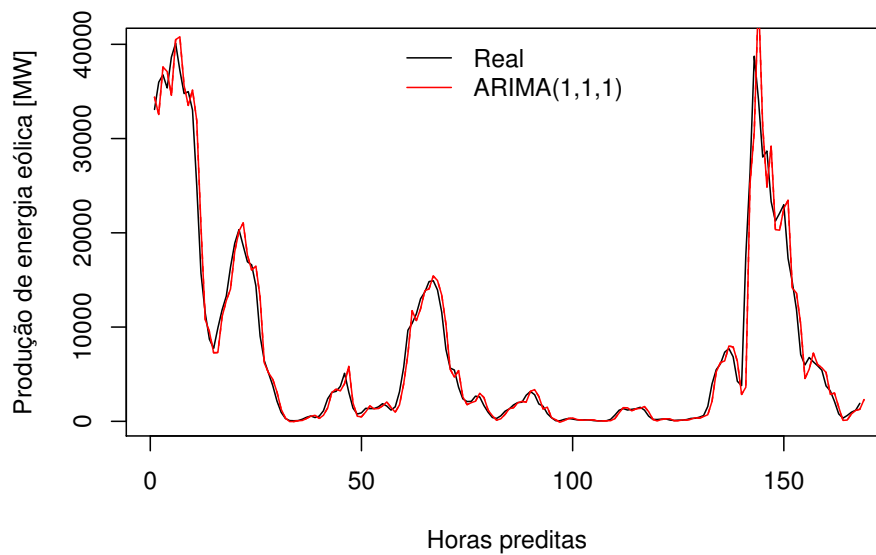


Fonte: Autor.

O $ARIMA(1,1,1)$ foi o selecionado dentre os parâmetros testados, dos quais foram o p entre 1 a 4, $d=1$, visto que foi preciso uma diferenciação para deixar a série estacionária,

e o q de 1 a 4. A escolha se deu por ser o modelo mais parcimonioso e ter um o $MAPE$ menor dentre os modelos testados. Porém, apesar de atender a maioria dos pressupostos, o gráfico de correlograma obteve mais de 5 % de lags significativos em 40. Todavia, mesmo sem estar no melhor dos seus ajustes o modelo estatístico tradicional obteve um desempenho muito expressivo, como pode ser visto na Figura 34. Assim como pode ser visto na Tabela 11, no qual o modelo alcançou um $MAPE$ próximo aos modelos inteligentes.

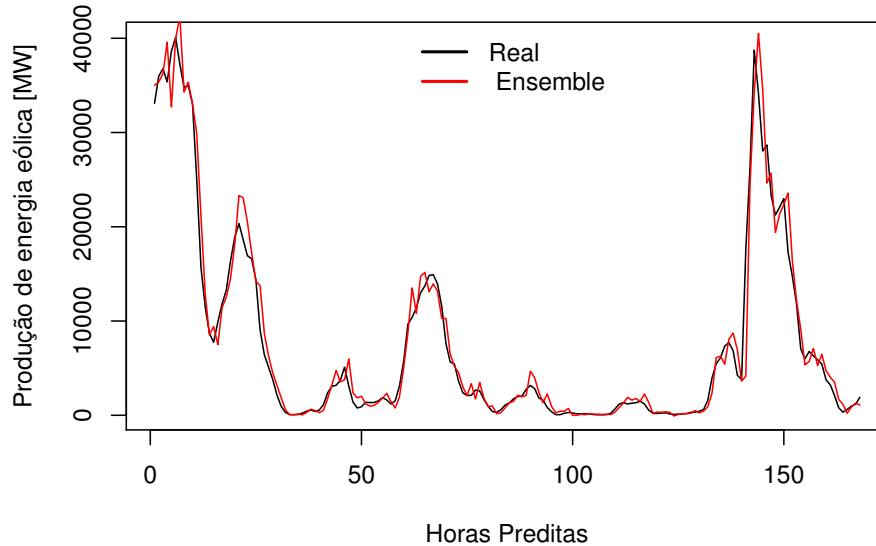
Figura 34 – Predição do modelo $ARIMA(1,1,1)$ na primeira estação.



Fonte: Autor.

Pode-se observar na Figura 35, que o *Ensemble XG* conseguiu, nos extremos, prever de uma forma mais precisa que os modelos individuais.

Figura 35 – Predição do modelo *Ensemble XG* na primeira estação.



Fonte: Autor.

Segundo a Tabela 11, o modelo que melhor se ajustou para a predição das 168 horas, verificando o *MAPE*, 1 passo a frente, na primeira estação, foi o *Random Forest*, seguido pelo *XGBoost*. Porém, o modelo *XGBoost* sem a seleção de variáveis foi extremamente mais efetiva do que o modelo *Random Forest* sem a seleção. O *Ensemble_XG*, combinador treinado pelo *XGBoost* e pelas predições dos modelos individuais, não superou os demais.

Tabela 11 – *MAPE* das 168 horas de predições na primeira estação.

Modelos	<i>MAPE</i> [%]
<i>XGBoost</i>	26
<i>XGBoost</i> sem seleção	32
<i>Random Forest</i>	22
<i>Random Forest</i> sem seleção	538
<i>ARIMA</i> (1,1,1)	43
<i>Ensemble_XG</i>	33

Fonte: Elaborada pelo autor.

Quanto ao *RMSE*, visto na Tabela 12, percebe-se que não houve uma diferença significativa entre os modelos *XGBoost* com e sem a seleção. Porém, com a seleção nos modelos *Random Forest* a melhora na predição foi evidente. Quanto ao *ensemble* proposto percebe-se uma melhoria relevante perante aos modelos individuais. Em geral, pode-se entender que o *Ensemble XG*, apesar de não ter sido o melhor modelo na média, nos momentos em que a produção

de energia eólica foram altos, o novo modelo errou menos. Ou seja, se o objetivo do modelo for não errar significativamente, o *Ensemble XG* é o mais interessante.

Tabela 12 – RMSE das 168 horas de predições na primeira estação.

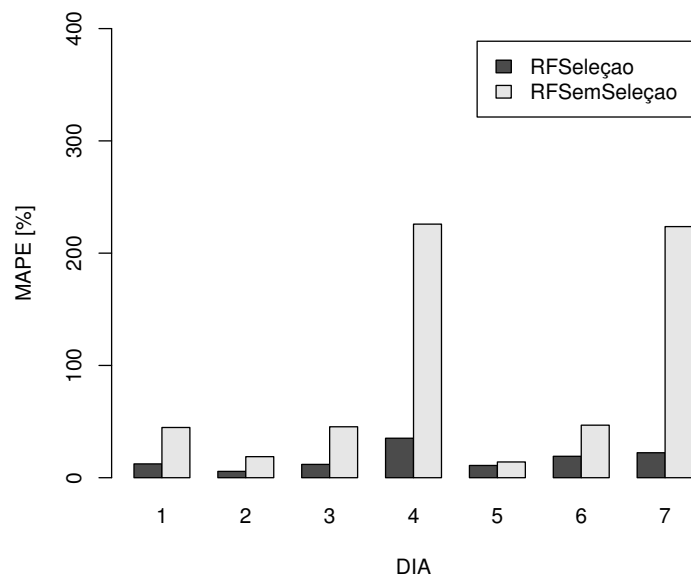
Modelos	RMSE [MW]
<i>XGBoost</i>	1952
<i>XGBoost</i> sem seleção	2051
<i>Random Forest</i>	1911,22
<i>Random Forest</i> sem seleção	5244
<i>ARIMA(1,1,1)</i>	2764,80
<i>Ensemble XG</i>	1497,08

Fonte: Elaborada pelo autor.

4.2.5 Resultados para segunda estação

A seleção de variáveis para os modelos *Random Forest*, referentes a segunda estação alvo, melhoraram significativamente o *MAPE*, como pode ser verificado na Figura 36.

Figura 36 – MAPE dos dias do *Random Forest* na segunda estação.

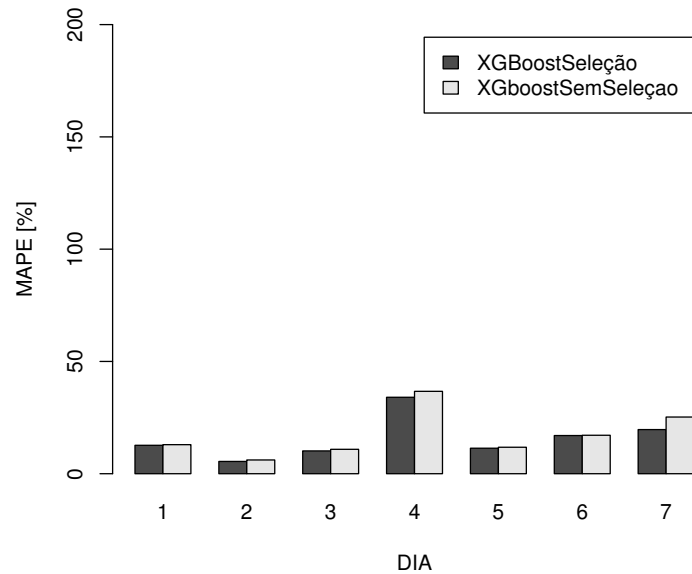


Fonte: Autor.

A seleção de variáveis para os modelos *XGBoost* referentes a segunda estação melhoraram, de acordo com o *MAPE*, como pode ser verificado na Figura 37. E, assim como na primeira

estação, o impacto não foi tão grande, como nos modelos *Random Forest*. Assim, se começa a perceber um padrão em relação aos modelos: O *XGBoost* consegue selecionar as melhores variáveis sem a necessidade dos métodos de seleção, diferente do *Random Forest*.

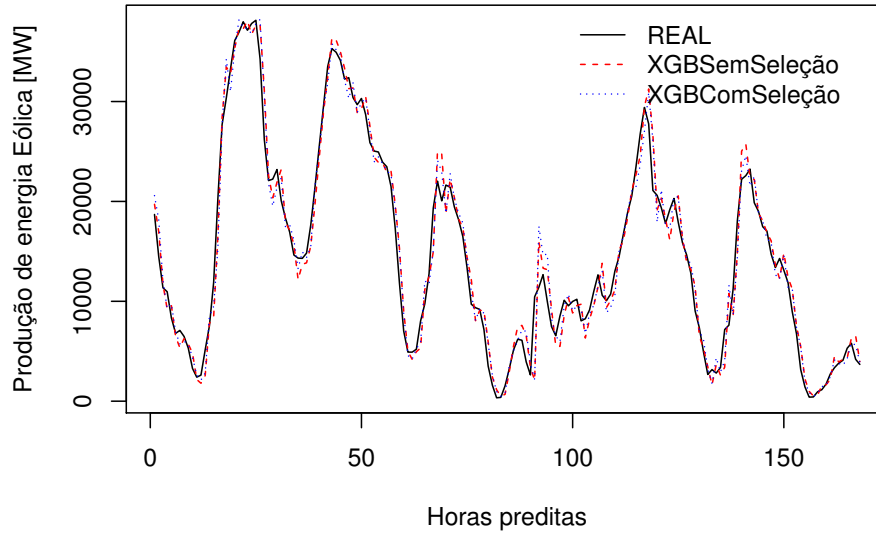
Figura 37 – MAPE dos dias do *XGBoost* na segunda estação.



Fonte: Autor.

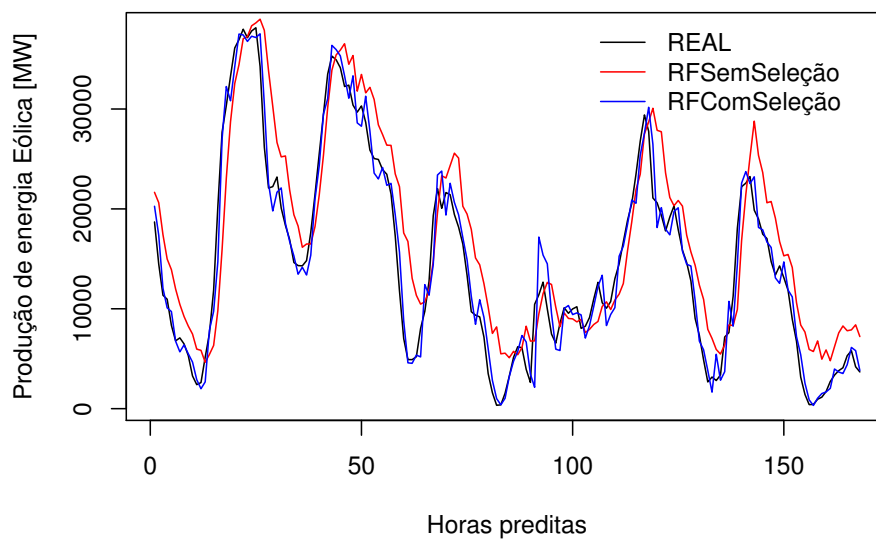
Deve ser ressaltado que nas Figuras 38 e 39 um padrão, assim como visto na primeira estação: sempre que existe um pico, os modelos *Random Forest* e *XGBoost*, supervalorizam suas predições, principalmente quando esta diferença de uma predição para outra é muito alta.

Figura 38 – Predição dos modelos *XGBoost* na segunda estação.



Fonte: Autor.

Figura 39 – Predição dos modelos *Random Forest* na segunda estação.

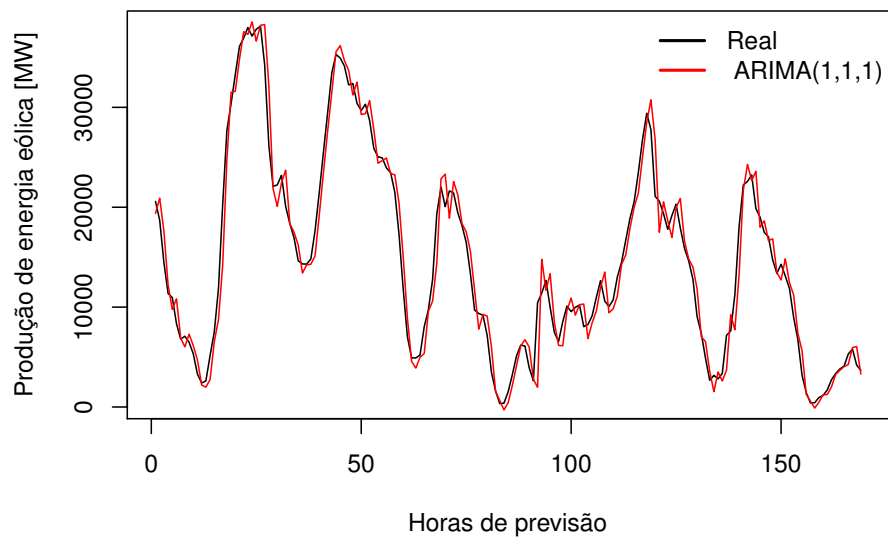


Fonte: Autor.

Como pode-se destacar na Figura 41, o modelo *Ensemble XG* se diferencia dos outros, principalmente, por conseguir prever melhor quanto aos extremos. O *ARIMA*, por sua vez,

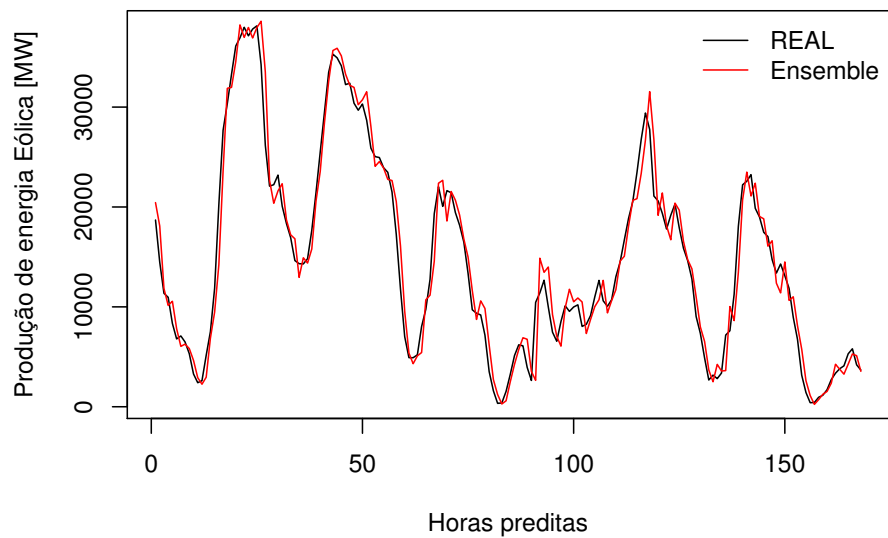
como pode ser visto na Figura 40, apresentou resultado parecido com dos modelos individuais, pecando nos mesmos pontos.

Figura 40 – Predição do $ARIMA(1,1,1)$ na segunda estação.



Fonte: Autor.

Figura 41 – Predição do *ensemble XG* na segunda estação.



Fonte: Autor.

Na Tabela 13 referente ao *MAPE* das 168 predições da segunda estação, o *XGBoost* foi melhor que os outros modelos individuais, porém vale ressaltar que os modelos *Random Forest*, apesar da melhora com a seleção, não foram tão significativas quanto a referente a primeira estação. Outro ponto importante é acerca dos modelos *XGBoost*, visto que o modelo sem a seleção manteve-se, em todas as predições, resultados relevantes. O *ARIMA(1,1,1)*, como na primeira estação, conseguiu um *MAPE* próximo aos modelos inteligentes. Quanto ao *ensemble* teve um desempenho melhor do que os modelos individuais.

Tabela 13 – *MAPE* das 168 horas de predições na segunda estação.

Modelos	<i>MAPE</i> [%]
<i>XGBoost</i>	15
<i>XGBoost</i> sem seleção	17
<i>Random Forest</i>	16
<i>Random Forest</i> sem seleção	88
<i>ARIMA(1,1,1)</i>	19
<i>Ensemble</i>	8

Fonte: Elaborada pelo autor.

Como visto na Tabela 14, os modelos *XGBoost* apresentaram *RMSE* extremamente próximos, porém, nesta estação o *RMSE* do modelo sem a seleção foi menor, o *Random Forest*, em contrapartida, assim como na primeira estação, obteve uma evolução significativa. O *ensemble XG*, além de ter obtido um *MAPE* menor, também obteve uma relevância maior no que se refere ao *RMSE* para com os modelos individuais.

Tabela 14 – *RMSE* das 168 horas de predições na segunda estação.

Modelos	<i>RMSE</i> [MW]
<i>XGBoost</i>	1900,27
<i>XGBoost</i> sem seleção	1881,39
<i>Random Forest</i>	1930,97
<i>Random Forest</i> sem seleção	4694,61
<i>ARIMA(1,1,1)</i>	1982,23
<i>Ensemble XG</i>	1028,69

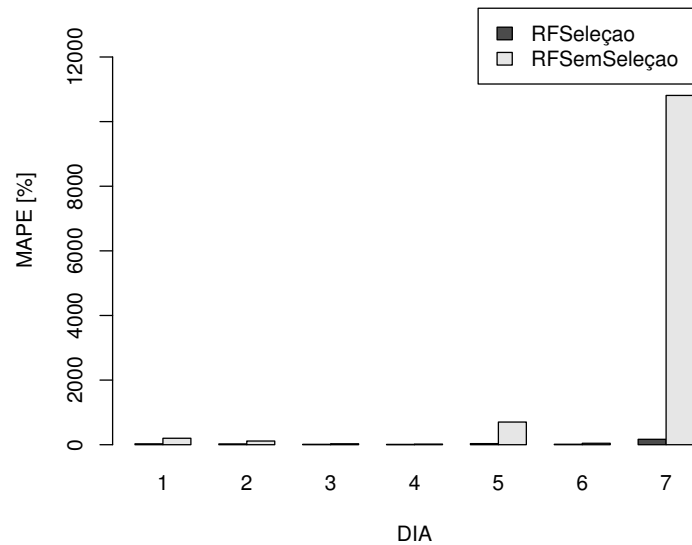
Fonte: Elaborada pelo autor.

4.2.6 Resultados para terceira estação

Nas predições referentes a terceira estação, a seleção de variáveis nos modelos *Random Forest*, assim como nos outros, obtiveram extrema relevância na melhora do *MAPE*, principal-

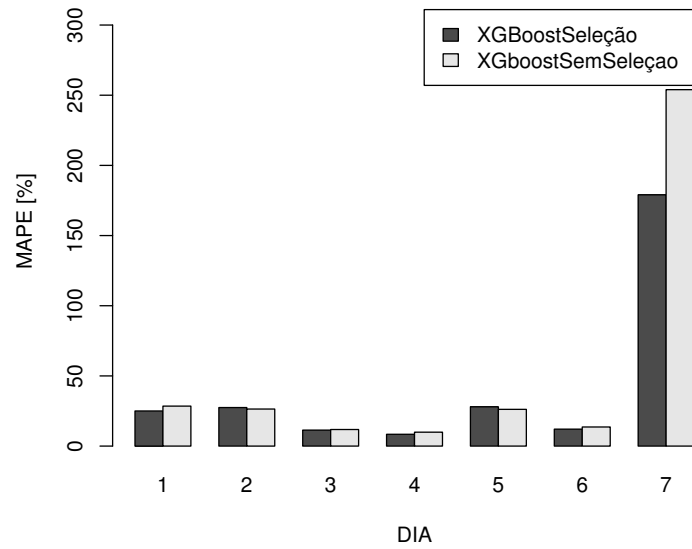
mente referente ao último dia de predição, visto que o *MAPE* do *Random Forest* sem a seleção foi acima dos demais. Com a seleção, o dia 7 também foi o maior, porém, em menor escala, como percebe-se na Figura 42.

Figura 42 – *MAPE* dos dias do *Random Forest* na terceira estação.



Fonte: Autor.

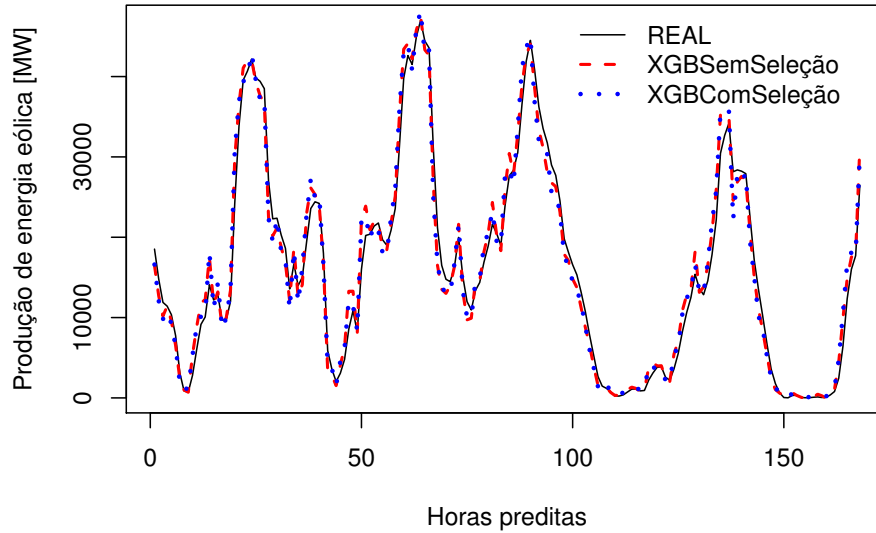
Quanto os modelos *XGBoost*, a maior diferença de *MAPE* também aconteceu no último dia de predição, com o modelo com a seleção de variáveis conseguindo responder melhor aos dados, de acordo com a Figura 43 .

Figura 43 – MAPE dos dias do *XGBoost* na terceira estação.

Fonte: Autor.

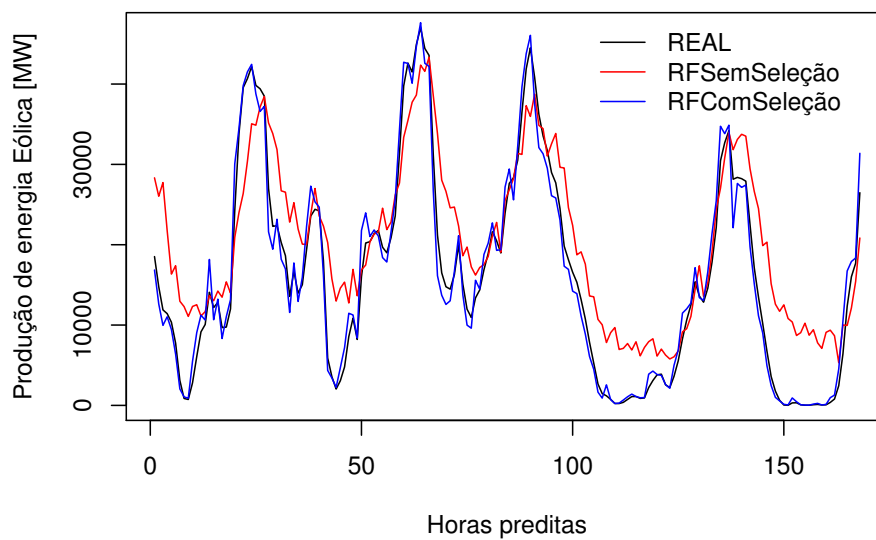
Como pode ser percebido nas Figuras 44 e 45, vale destacar a última predição da hora (168), que todos os modelos a superestimaram. Porém, tal fato não foi exclusivo para esta hora, como pode-se observar que, na maioria destes extremos, todos os modelos superestimam suas predições.

Figura 44 – Predição dos modelos *XGBoost* na terceira estação.



Fonte: Autor.

Figura 45 – Predição dos modelos *Random Forest* na terceira estação.

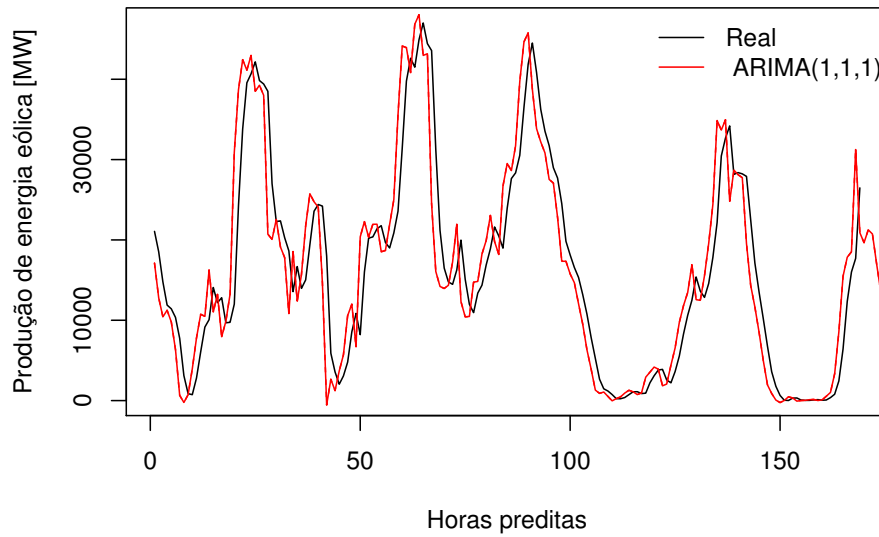


Fonte: Autor.

O modelo *ARIMA(1,1,1)*, Figura 46, assim como os outros modelos, não conseguiu prever com uma acurácia alta os últimos dados da terceira estação, porém, não obteve tantas

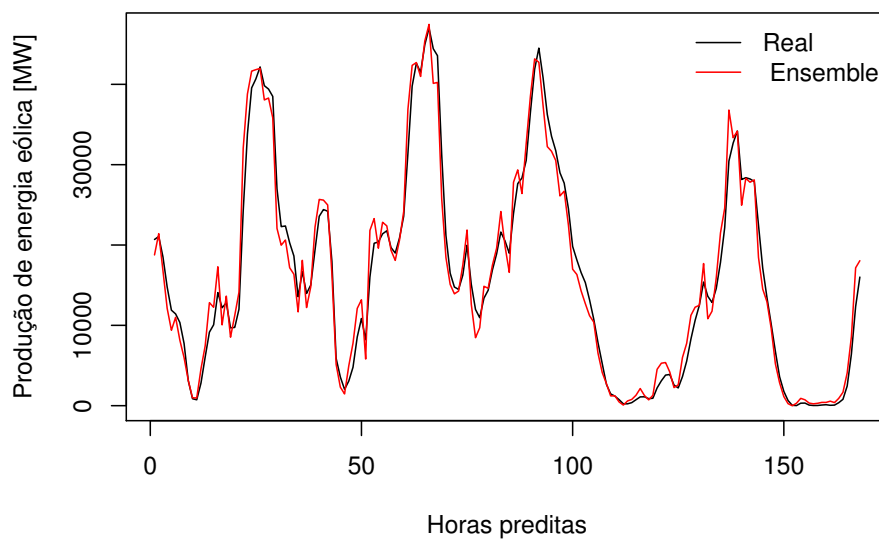
dificuldades nos momentos de pico.

Figura 46 – Predição do modelo $ARIMA(1,1,1)$ na terceira estação.



Fonte: Autor.

Figura 47 – Predição do modelo *ensemble* na terceira estação.



Fonte: Autor.

Na Tabela 15 consta-se que os modelos *Random Forest* e *XGBoost* obtiveram o mesmo

desempenho em relação ao *MAPE* nas 168 predições referentes a terceira estação. O modelo *XGBoost* sem a seleção obteve um resultado relevante em comparação com os dois melhores modelos, porém, o *Random Forest* sem a seleção obteve um resultado bem inferior, visto que seu *MAPE* chegou a 16,30. Quanto ao *ARIMA*, manteve desempenho, assim, como em outras estações, resultados comparáveis aos modelos inteligentes, porém, sempre um pouco abaixo.

Tabela 15 – *MAPE* das 168 horas de predições na terceira estação.

Modelos	MAPE [%]
<i>XGBoost</i>	41
<i>XGBoost</i> sem seleção	52
<i>Random Forest</i>	41
<i>Random Forest</i> sem seleção	1630
<i>ARIMA</i> (1,1,1)	48
<i>Ensemble XG</i>	43

Fonte: Elaborada pelo autor.

Como visto na Tabela 16, o *RMSE* para com os modelos *XGBoost* aumenta após a seleção de variáveis, porém, não é um aumento significativo. Quanto aos modelos *Random Forest* a queda do *RMSE* é extremamente alta e a melhora na predição também. Já com o modelo *ensemble*, Figura 47, pode-se entender que, apesar de não ter acertado mais que os outros modelos individuais na média, quando errou, errou por menos.

Tabela 16 – *RMSE* das 168 horas de predição na terceira estação.

Modelos	RMSE [MW]
<i>XGBoost</i>	2994,92
<i>XGBoost</i> sem seleção	2986,33
<i>Random Forest</i>	3027,18
<i>Random Forest</i> sem seleção	8138,07
<i>ARIMA</i> (1,1,1)	3019,12
<i>Ensemble XG</i>	2077,47

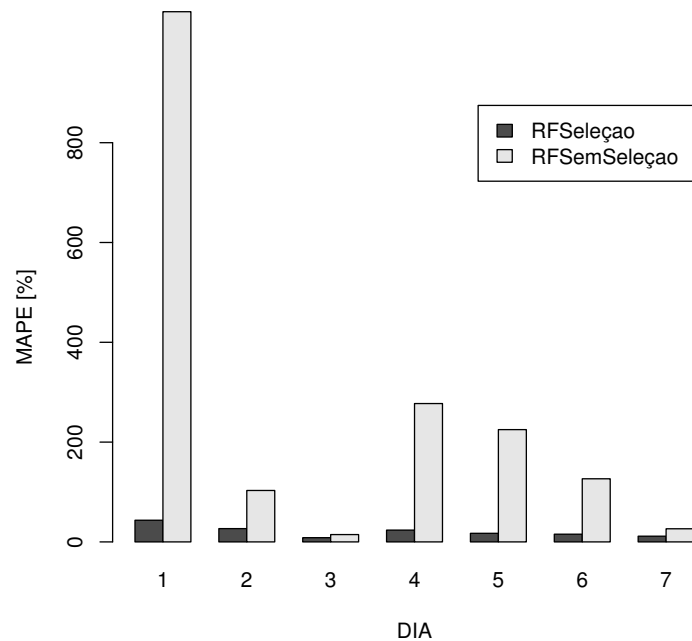
Fonte: Elaborada pelo autor.

4.2.7 Resultados para quarta estação

Como visto na Figura 48, a seleção de variáveis, principalmente no primeiro dia de predição, obteve um papel importante no que se refere ao *MAPE*. É interessante perceber que, apesar de ter um *MAPE* maior, em média, o *MAPE* dos modelos com a seleção parecem pro-

porcionais aos sem a seleção, ou seja, quando os modelos sem a seleção erram mais em um dia específico, os modelos com a seleção também errarão mais nesse dia.

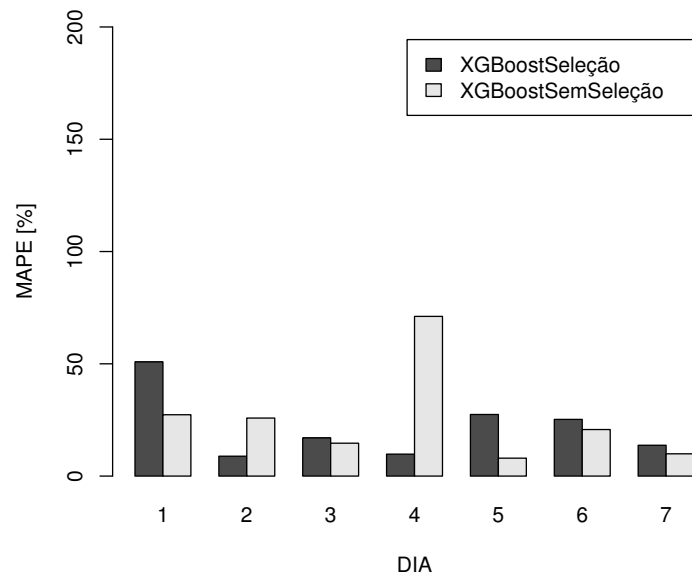
Figura 48 – MAPE dos dias do *Random Forest* na quarta estação.



Fonte: Autor.

Quanto a seleção aplicada aos modelos *XGBoost* para a quarta estação, percebe-se, segundo a Figura 49, que como no modelo *Random Forest*, o mesmo também apresentou seu maior *MAPE* no primeiro dia de predição, e com a seleção de variáveis a diminuição do *MAPE* foi interessante. Como nas outras estações, a seleção de variáveis não foi tão relevante quanto nos modelos *Random Forest*, porém, foi observado uma melhora.

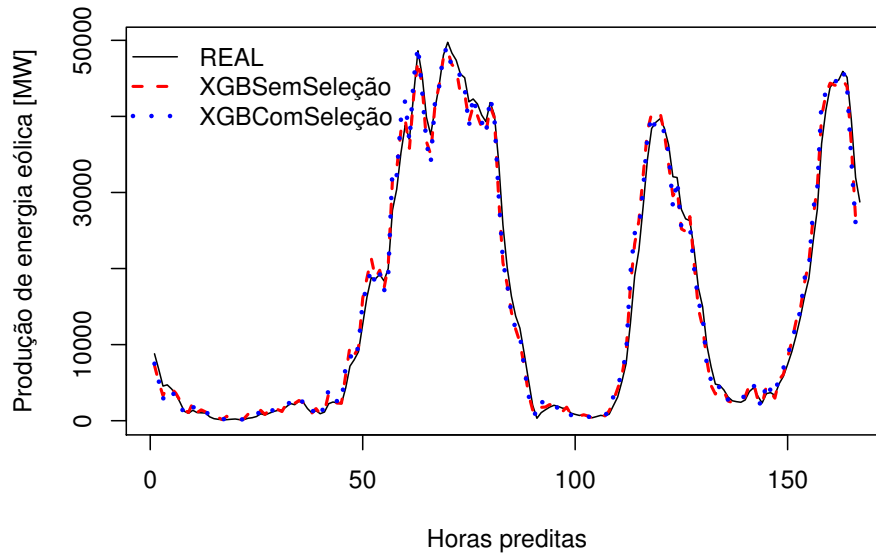
Figura 49 – MAPE dos dias do XGBoost na quarta estação.



Fonte: Autor.

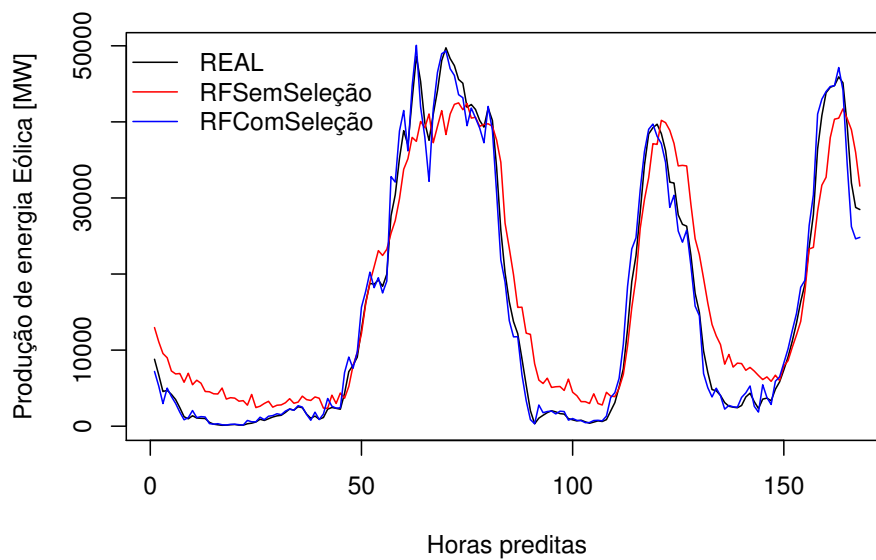
Assim como nas outras predições, é visto nas Figuras 50 e 51 a supervalorização dos modelos *Random Forest* e *XGBoost* nos picos. Vale destacar também que o modelo *Random Forest* sem a seleção de variáveis não conseguiu interpretar bem os dados e foi muito mal em comparação com os demais.

Figura 50 – Predição dos modelos *XGBoost* na quarta estação.



Fonte: Autor.

Figura 51 – Predição dos modelos *Random Forest* na quarta estação.

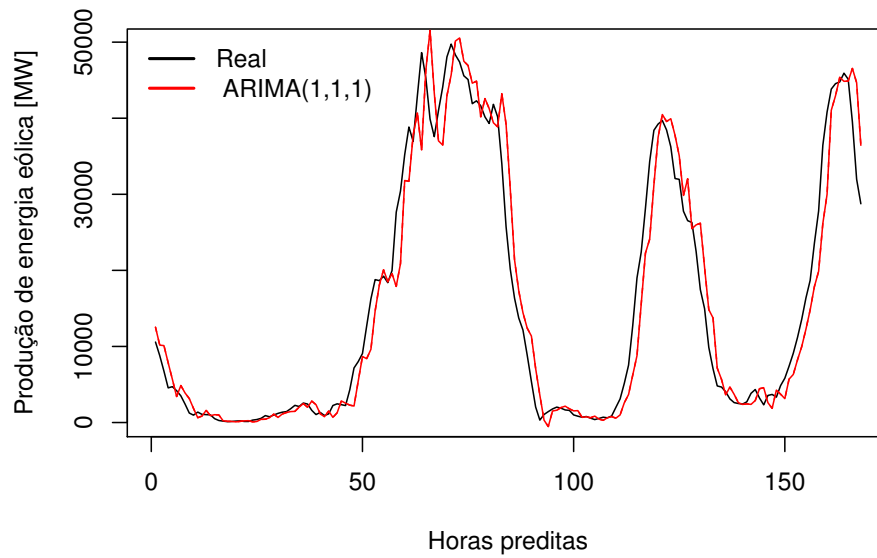


Fonte: Autor.

O *Ensemble*, em contramão dos modelos individuais, não superestimava os extremos dos modelos. Na realidade, nesses picos o modelo *ensemble* conseguia se sair muito bem, conforme

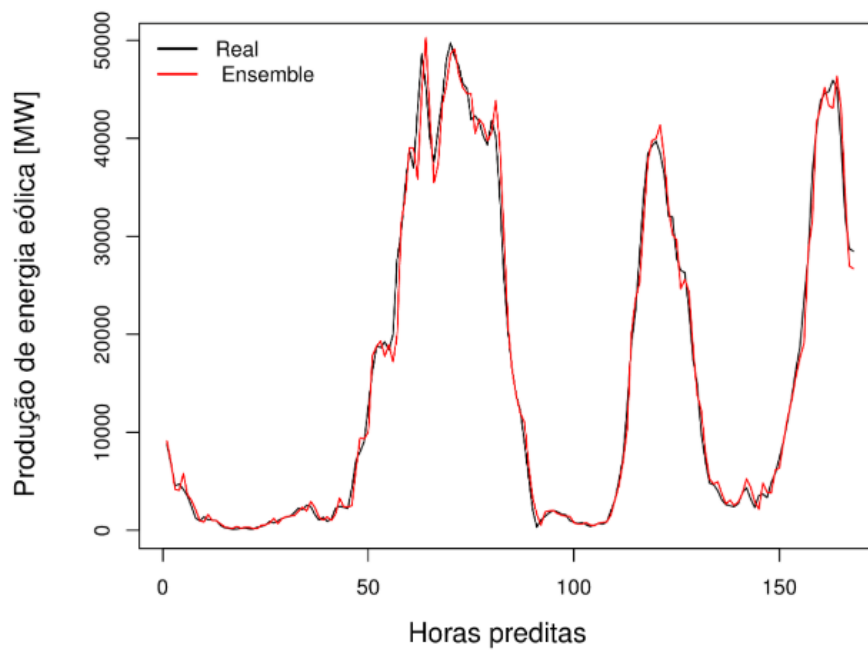
pode ser visto também na Figura 53, o que merece ser destacado visto a enorme dificuldade dos modelos individuais para estes casos.

Figura 52 – Predição do modelo *ARIMA(1,1,1)* na quarta estação.



Fonte: Autor.

Figura 53 – Predição do modelo *Ensemble XG* na quarta estação.



Fonte: Autor.

Como pode ser visto na Tabela 17, o melhor modelo individual, segundo o *MAPE*, foi o *Random Forest* com a seleção, seguido do *XGBoost* com a seleção. O *XGBoost* sem a seleção obteve um resultado muito próximo aos dois melhores. Já o *Random Forest* sem a seleção apresentou um resultado menos significativo em relação aos demais. Quanto ao *ensemble*, percebe-se que o mesmo foi melhor em relação ao *MAPE*, do que todos os modelos individuais apresentados.

Tabela 17 – *MAPE* das 168 horas de predições na quarta estação.

Modelos	<i>MAPE</i> [%]
<i>XGBoost</i>	22
<i>XGBoost</i> sem seleção	25
<i>Random Forest</i>	20
<i>Random Forest</i> sem seleção	262
<i>ARIMA</i> (1,1,1)	48
<i>Ensemble XG</i>	18

Fonte: Elaborada pelo autor.

Na quarta estação, como pode ser verificado na Tabela 18, os modelos também se beneficiaram com a seleção de variáveis, em relação ao *RMSE*. Tanto o *XGBoost* como o *Random Forest* apresentaram uma grande diminuição. Em relação ao *ensemble* proposto, é visto que obteve um desempenho melhor comparado aos modelos individuais, conforme visto nas Tabelas 17 e 18.

Tabela 18 – *RMSE* das 168 predições na quarta estação.

Modelos	<i>RMSE</i> [MW]
<i>XGBoost</i>	1810,05
<i>XGBoost</i> sem seleção	4000,89
<i>Random Forest</i>	2164,38
<i>Random Forest</i> sem seleção	5015,69
<i>ARIMA</i> (1,1,1)	2960,93
<i>Ensemble XG</i>	1658,65

Fonte: Elaborada pelo autor.

Percebe-se, analisando os resultados das 4 estações do ano, que na segunda estação alvo foi a estação no qual os modelos obtiveram os melhores resultados, enquanto que na terceira estação alvo os modelos apresentaram um *MAPE* e *RMSE*, em média, maior, além do maior erro de um modelo individual ser de tal estação (*Random Forest* sem seleção, que chegou a ter um *MAPE* de 1630).

4.2.8 Análise geral dos modelos

Em relação aos resultados dos modelos individuais, pode-se separar a análise em duas etapas: Pelas métricas e pelo tempo de desempenho. Além disso, também é importante analisar o impacto da seleção de variável nos mesmos.

Quanto a esta seleção, percebe-se que o modelo *Random Forest* foi o mais beneficiado, visto que o modelo com diversas variáveis não conseguiu se sair bem nas predições, devido ao seu método de aproximação. No entanto, no caso do *XGBoost*, a seleção não teve um impacto significativo, ou seja, o mesmo foi capaz de dar os pesos ideais para as diversas variáveis, mesmo sendo um número excessivo, 336 covariáveis.

No que se refere ao *MAPE* e *RMSE*, nos modelos individuais, pode se colocar o *XGBoost* e o *Random Forest* no mesmo patamar, visto que o desempenho destes modelos foram muito parecidos. Porém, vale destacar que o *XGBoost*, em relação ao tempo de execução, foi mais eficiente. O *ARIMA*, um método mais tradicional, apesar de ter tido um desempenho pior, seu resultado foi relevante, justificando o motivo de ser um método tão usual.

Quanto a combinação dos modelos, o *Ensemble XG*, obteve um resultado bem relevante e supriu a principal deficiência dos modelos individuais: a superestimação nos momentos de extremos. Este fato direcionou um impacto relevante, principalmente, em relação ao *RMSE*. Porém, quanto ao *MAPE*, o novo modelo proposto não foi o melhor em todas as estações, mas apresentou bom resultado, superando os modelos individuais em duas estações.

5 CONCLUSÕES E TRABALHOS FUTUROS

O trabalho teve como objetivo fazer a predição de uma semana, em escala horária, da produção eólica em quatro diferentes períodos do ano, utilizando os modelos *Random Forest*, *XGBoost* e o *ARIMA*, além de aplicar um modelo de seleção de variáveis nos métodos inteligentes, propondo, por fim, um método *ensemble* com os modelos individuais.

O modelo *ensemble* proposto tinha como ideia acrescentar os resultados individuais de cada modelo. Para o treino do novo método, utilizou-se 700 horas previstas antes da semana alvo. Ou seja, se fosse prever a produção no tempo t , fez-se a predição de $t - 700$ até t com os modelos individuais e usou essa nova base como treino para o novo modelo.

A partir dos resultados obtidos, percebe-se que, o modelo *Random Forest* foi o mais beneficiado, proporcionalmente, com o algoritmo de seleção utilizado. O modelo *XGBoost*, devido ao seu método de aproximação, conseguiu, mesmo sem a seleção, ter um resultado muito positivo. O que deve ser destacado, visto que este algoritmo consegue ter um bom desempenho e selecionar as principais variáveis para a predição, mesmo em um banco de dados com diversas variáveis.

Em geral, os modelos *XGBoost* e *Random Forest* com a aplicação da seleção tiveram desempenho muito parecido, com pequena vantagem no *MAPE* para o *Random Forest* e pequena vantagem no *RMSE* para o *XGBoost*. Quanto ao *ARIMA*, apresentou um desempenho ligeiramente pior, porém, melhor que o *Random Forest* sem a seleção.

Já o *ensemble* proposto obteve um desempenho, segundo as métricas de avaliação utilizadas, melhor que os modelos desenvolvidos individualmente, além de minimizar os erros de supervalorização nas predições dos extremos apresentadas por estes modelos individuais, com exceção do *MAPE* em duas ocasiões.

Quanto os resultados nas estações, percebe-se que as semanas com maior variabilidade foram as mais complicadas para os modelos no que se refere as métricas de avaliação utilizadas.

Com isso, vale destacar que o método de seleção aplicado com as variáveis mais relevantes do *Mutual Information* e *VarImportance* apresentaram, em geral, um bom desempenho para a seleção, pois, além melhorarem os resultados quando foram aplicados, diminuíram, e muito, o tempo de suas execuções, tornando os novos modelos mais eficazes e eficientes.

Como sugestões para trabalhos futuros utilizar os métodos abordados no trabalho com alguma fazenda eólica no Brasil, além de acrescentar outras variáveis físicas, como a pressão

atmosférica, direção do vento e temperatura. Porém, também seria interessante aplicar os modelos em outros banco de dados, para conferir acerca da confiabilidade dos modelos e do método *ensemble*.

REFERÊNCIAS

- ACKERMANN, T. **Wind power in power systems**. [S.l.]: John Wiley & Sons, 2005.
- ACKERMANN, T.; SÖDER, L. An overview of wind energy-status 2002. **Renewable and sustainable energy reviews**, Elsevier, v. 6, n. 1-2, p. 67–127, 2002.
- AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. ANEEL. **Matriz de energia elétrica**. 2016. Disponível em: <<http://www.aneel.gov.br/>>. Acesso em: 12 maio 2018.
- AMJADY, N.; DARAEPOUR, A. Design of input vector for day-ahead price forecasting of electricity markets. **Expert Systems with Applications**, Elsevier, v. 36, n. 10, p. 12281–12294, 2009.
- AMJADY, N.; DARAEPOUR, A.; KEYNIA, F. Day-ahead electricity price forecasting by modified relief algorithm and hybrid neural network. **IET generation, transmission & distribution**, IET, v. 4, n. 3, p. 432–444, 2010.
- AMJADY, N.; KEYNIA, F. Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm. **IEEE Transactions on Power Systems**, IEEE, v. 24, n. 1, p. 306–318, 2009.
- ANDRADE, R. O. B. d.; TACHIZAWA, T.; CARVALHO, A. B. d. **Gestão ambiental: enfoque estratégico aplicado ao desenvolvimento sustentável**. [S.l.]: São Paulo: Makron Books, 2000. v. 232.
- ARBOIT, N. K. S. et al. Potencialidade de utilização da energia geotérmica no Brasil—uma revisão de literatura. **Revista do Departamento de Geografia**, v. 26, p. 155–168, 2013.
- ASSOCIAÇÃO BRASILEIRA DE ENERGIA EÓLICA. ABEE. **Energia Eólica**. 2018. Disponível em: <<https://www.abeolica.org.br/>>. Acesso em: 13 nov. 2017.
- BORGES, F. H.; TACHIBANA, W. K. A evolução da preocupação ambiental e seus reflexos no ambiente dos negócios: uma abordagem histórica. **XXV Encontro Nac. de Eng. De Produção—Porto Alegre**, 2005.
- BREIMAN, L. Bagging predictors. **Machine Learning**, v. 24, n. 2, p. 123–140, 1996. ISSN 1573-0565. Disponível em: <<http://dx.doi.org/10.1023/A:1018054314350>>.
- _____. Random forests. **Machine learning**, Springer, v. 45, n. 1, p. 5–32, 2001.
- BREIMAN, L. et al. **Classification and Regression Trees**. [S.l.]: California, Wadsworth. Inc, 1984.
- BRUHN, K.; LORENSSON, S.; SVENSSON, J. **Development of learning material to wind power courses**. 2009. 180 f. Dissertação (Mestrado em Engenharia Elétrica) — Chalmers University of Technology, Gothenburg, 2009.

CALDEIRA, A. M.; SOUZA, R. C.; MACHADO, M. A. S. Identificação automática das ordens dos modelos garch utilizando redes neurais. **Engevista**, v. 11, n. 2, 2010.

CARDOSO, E. d. M. et al. Apostila educativa: Energia nuclear. **Comissão Nacional de Energia Nuclear (CNEN)**, 2001.

CHANG, G. et al. A hybrid model for forecasting wind speed and wind power generation. In: IEEE. **Power and Energy Society General Meeting (PESGM), 2016**. [S.l.], 2016. p. 1–5.

CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: ACM. **Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining**. [S.l.], 2016. p. 785–794.

CUSTÓDIO, R. Energia eólica para produção de energia elétrica. rio de janeiro: Synergia. **Acta**, 2013.

DARAEPOUR, A.; ECHEVERRI, D. P. Day-ahead wind speed prediction by a neural network-based model. In: IEEE. **Innovative Smart Grid Technologies Conference (ISGT), 2014 IEEE PES**. [S.l.], 2014. p. 1–5.

DÍAZ-URIARTE, R.; ANDRES, S. A. D. Gene selection and classification of microarray data using random forest. **BMC bioinformatics**, BioMed Central, v. 7, n. 1, p. 3, 2006.

DODGE, D. M. Illustrated history of wind power development. “2002. **Darrell Dodge and TelosNet Web Development**, 2006.

ELDALI, F. A. et al. Employing arima models to improve wind power forecasts: A case study in ercot. In: IEEE. **North American Power Symposium (NAPS), 2016**. [S.l.], 2016. p. 1–6.

FERREIRA, A. A. B. **Metodologia de previsão do potencial eólico de curto prazo para planejamento da operação de sistemas elétricos**. 2012. 97 f. Dissertação (Mestrado em Engenharia) — Universidade Federal do Pampa, Alegrete, 2012.

FIGUEIREDO, M. et al. Gás natural: potencialidades de utilização no brasil. **Revista Eletrônica em Gestão, Educação e Tecnologia Ambiental (Fechada para submissões por tempo indeterminado)**, v. 10, n. 10, p. 2251–2265, 2013.

FLEMING, F. P. **Avaliação do potencial de energias oceânicas no Brasil**. Dissertação (Mestrado em Planejamento Estratégico, school=Universidade Federal do Rio de Janeiro, address=Rio de Janeiro, 2012, pages=100, pagename = f., year=2012).

FOLEY, A. M. et al. Current methods and advances in forecasting of wind power generation. **Renewable Energy**, Elsevier, v. 37, n. 1, p. 1–8, 2012.

FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. **Annals of statistics**, JSTOR, p. 1189–1232, 2001.

FUGON, L.; JUBAN, J.; KARINIOTAKIS, G. Data mining for wind power forecasting. In:

EWEC. **European Wind Energy Conference & Exhibition EWEC 2008**. [S.l.], 2008. p. 6–pages.

GLOBAL WIND ENERGY COUNCIL. GWEC. **Wind Energy**. 2018. Disponível em: <<https://www.gwec.net>>. Acesso em: 13 nov. 2017.

HOLTTINEN, H.; MIETTINEN, J.; SILLANPÄÄ, S. Wind power forecasting accuracy and uncertainty in finland. **Espoo, VTT**, 2013.

HONG, D. et al. An indirect short-term wind power forecast approach with multi-variable inputs. In: IEEE. **Innovative Smart Grid Technologies-Asia (ISGT-Asia), 2016 IEEE**. [S.l.], 2016. p. 793–798.

JUNQUEIRA, M. Brasil é líder em projetos de carbono. **Gazeta Mercantil**, 2006.

JURADO, S. et al. Short-term electric load forecasting using computational intelligence methods. In: IEEE. **Fuzzy Systems (FUZZ), 2013 IEEE International Conference on**. [S.l.], 2013. p. 1–8.

JURSA, R. et al. Wind power prediction with different artificial intelligence models. In: **Proceedings of the 2007 European Wind Energy Conference and Exhibition (EWEC)**. [S.l.: s.n.], 2007. p. 1–10.

KALDELLIS, J. K.; ZAFIRAKIS, D. The wind energy (r) evolution: A short review of a long history. **Renewable Energy**, Elsevier, v. 36, n. 7, p. 1887–1901, 2011.

KOHAVI, R.; JOHN, G. H. Wrappers for feature subset selection. **Artificial intelligence**, Elsevier, v. 97, n. 1-2, p. 273–324, 1997.

KOTO, S. M. **Fontes renováveis para a produção de energia elétrica no contexto do Mecanismo de Desenvolvimento Limpo**: estudo de caso de duas fontes-eólica e fotovoltaica. 2009. 60 f. Dissertação (Especialista em Gestão Ambiental e Negócios no Setor Energético) — Universidade de São Paulo, São Paulo, 2009.

KUSIAK, A.; ZHENG, H.; SONG, Z. Wind farm power prediction: a data-mining approach. **Wind Energy**, Wiley Online Library, v. 12, n. 3, p. 275–293, 2009.

LAHOUAR, A.; SLAMA, J. B. H. Day-ahead load forecast using random forest and expert input selection. **Energy Conversion and Management**, Elsevier, v. 103, p. 1040–1051, 2015.

LEI, M. et al. A review on the forecasting of wind speed and generated power. **Renewable and Sustainable Energy Reviews**, Elsevier, v. 13, n. 4, p. 915–920, 2009.

LI, S.; WANG, P.; GOEL, L. Wind power forecasting using neural network ensembles with feature selection. **IEEE Transactions on Sustainable Energy**, IEEE, v. 6, n. 4, p. 1447–1456, 2015.

LIMA, J. M. **Modelo meteorológico-estocástico para previsão da geração de energia**

- eólica**. 2016. 199 f. Tese (Doutorado em Engenharia de Recursos Hídricos) — Universidade Federal do Paraná, Curitiba, 2016.
- LIU, H. et al. A hybrid statistical method to predict wind speed and wind power. **Renewable energy**, Elsevier, v. 35, n. 8, p. 1857–1861, 2010.
- LIU, W.; PARK, I.; PRINCIPE, J. C. An information theoretic approach of designing sparse kernel adaptive filters. **IEEE Transactions on Neural Networks**, IEEE, v. 20, n. 12, p. 1950–1961, 2009.
- LUND, J. W.; FREESTON, D. H.; BOYD, T. L. Direct utilization of geothermal energy 2010 worldwide review. **Geothermics**, Elsevier, v. 40, n. 3, p. 159–180, 2011.
- MEBRATU, D. Sustainability and sustainable development: historical and conceptual review. **Environmental impact assessment review**, Elsevier, v. 18, n. 6, p. 493–520, 1998.
- MONTEIRO, C. et al. **Wind power forecasting: state-of-the-art 2009**. [S.l.], 2009.
- MOREIRA, P. H. V. et al. Aplicação de séries temporais na previsão da média mensal da taxa de câmbio do real para o dólar comercial de compra usando o modelo de holt. 2010.
- MORETTIN, P.; TOLOI, C. **Análise de Séries Temporais**. [S.l.]: 2. ed. São Paulo: Editora E. Blücher-Associação Brasileira de Estatística, 2006.
- OPERADOR NACIONAL DO SISTEMA ELTRICO. ONS. **Sistema elétrico**. 2018. Disponível em: <<http://ons.org.br/>>. Acesso em: 12 maio 2018.
- ORSOLON, M. Crédito de carbono. **Potência**, p. 16–26, 2006.
- RÜNCOS, F. et al. Geração de energia eólica—tecnologias atuais e futuras. **WEG Maquinas—GRUCAD-EEL-TET-UFSC**, 2000.
- RUPPERT, D. **Statistics and data analysis for financial engineering**. [S.l.]: Springer, 2011. v. 13.
- SFETSOS, A. A comparison of various forecasting techniques applied to mean hourly wind speed time series. **Renewable energy**, Elsevier, v. 21, n. 1, p. 23–35, 2000.
- SILVA, A. R. **Energia eólica em alto mar: distribuição dos recursos e complementaridade hídrica**. 2015. 139 f. Tese (Doutorado em Ciências Climáticas) — Universidade Federal do Rio Grande do Norte, Natal, 2015.
- SOERENSEN, H. C.; WEINSTEIN, A. Ocean energy: position paper for ipcc. In: CITESEER. **IPCC Scoping Meeting on Renewable Energy Sources—Proceedings**. [S.l.], 2008. p. 93–102.
- TACONELI, C. A. **Árvores de classificação multivariadas fundamentadas em coeficientes de dissimilaridade e entropia**. 2008. Tese (Doutorado) — Escola Superior de Agricultura

“Luiz de Queiroz.

THOMAS, R. L.; SAVINO, J. M. Status of wind-energy conversion. 1973.

THORNTON, H. E. et al. The relationship between wind power, electricity demand and winter weather patterns in great britain. **Environmental Research Letters**, IOP Publishing, v. 12, n. 6, p. 064017, 2017.

TOLMASQUIM, M. As origens da crise energética brasileira. **Ambiente & sociedade**, SciELO Brasil, n. 6-7, p. 179–183, 2000.

TUBINO, D. Administração dos estoques. **Manual de planejamento e controle da produção.**, São Paulo: Atlas, p. 103–145, 2000.

VIEIRA, P. L. et al. Gás natural: benefícios ambientais no estado da bahia. In: CONGRESSO BRASILEIRO DE PD EM PETRÓLEO E GÁS, 3, 2004. SALVADOR. **Anais.** [S.l.], 2005. p. 1–6.

VITERBO, J. C. **Geração de energia elétrica a partir da fonte eólica offshore.** 2008. 168 f. Dissertação (Mestrado em Engenharia) — Universidade de São Paulo, São Paulo, 2008.

WERNER, L.; RIBEIRO, J. L. D. Previsão de demanda: uma aplicação dos modelos box-jenkins na área de assistência técnica de computadores pessoais. **Gestão e produção.**, SciELO Brasil, v. 10, n. 1 (abr. 2003), p. 47–67, 2003.

YANG, L. et al. Support-vector-machine-enhanced markov model for short-term wind power forecast. **IEEE Transactions on Sustainable Energy**, IEEE, v. 6, n. 3, p. 791–799, 2015.

YU, L.; LIU, H. Efficient feature selection via analysis of relevance and redundancy. **Journal of machine learning research**, v. 5, n. Oct, p. 1205–1224, 2004.