



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE BACHARELADO EM ESTATÍSTICA**

PALLOMA KATTYANNE DE OLIVEIRA SILVA

**UTILIZAÇÃO DO MÉTODO DE REGRESSÃO LINEAR EM DADOS
OBSERVADOS E ESTIMADOS DA VELOCIDADE DO VENTO NA REGIÃO DO
SUBMÉDIO RIO SÃO FRANCISCO**

**CAMPINA GRANDE – PB
DEZEMBRO DE 2019**

PALLOMA KATTYANNE DE OLIVEIRA SILVA

**UTILIZAÇÃO DO MÉTODO DE REGRESSÃO LINEAR EM DADOS
OBSERVADOS E ESTIMADOS DA VELOCIDADE DO VENTO NA REGIÃO DO
SUBMÉDIO RIO SÃO FRANCISCO**

Trabalho de Conclusão de Curso (Artigo) apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Orientador: Gustavo Henrique Esteves
Coorientador: Pollyanna Kelly de Oliveira Silva

**CAMPINA GRANDE - PB
DEZEMBRO DE 2019**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

S586u Silva, Palloma Kattyane de Oliveira.

Utilização do método de regressão linear em dados observados e estimados da velocidade do vento na região do submédio Rio São Francisco [manuscrito] / Palloma Kattyane de Oliveira Silva. - 2019.

21 p.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2019.

"Orientação : Prof. Dr. Gustavo Henrique Esteves, Coordenação do Curso de Estatística - CCT."

"Coorientação: Profa. Dra. Pollyanna Kelly de Oliveira Silva, Coordenação do Curso de Estatística - CCT."

1. Projeto SONDA. 2. Regressão linear. 3. Recursos eólicos. 4. Dados de reanálises. I. Título

21. ed. CDD 519.5


PALLOMA KATTYANNE DE OLIVEIRA SILVA

UTILIZAÇÃO DO MÉTODO DE REGRESSÃO LINEAR EM DADOS OBSERVADOS
E ESTIMADOS DA VELOCIDADE DO VENTO NA REGIÃO DO SUBMÉDIO RIO
SÃO FRANCISCO

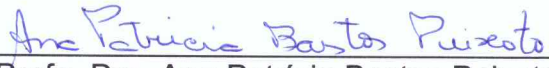
Trabalho de Conclusão de Curso (Artigo) apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Trabalho aprovado em 02 de dezembro de 2019.

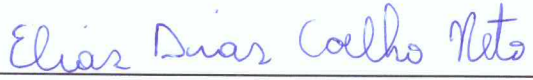
BANCA EXAMINADORA



Prof. Gustavo Henrique Esteves (Orientador)
Universidade Estadual da Paraíba (UEPB)



Profa. Dra. Ana Patrícia Bastos Peixoto
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Elias Dias Coelho Neto
Universidade Estadual da Paraíba (UEPB)

Acima de tudo, agradeço a Deus por mais esta realização.

Dedico a minha família que, com muito carinho e apoio, não mediram esforços para que eu chegasse até esta etapa de minha vida.

“Existem dois tipos de estatísticas: as que
você lê e as que você faz”

(Rex Stout)

LISTA DE ILUSTRAÇÕES

- Figura 1** – Gráficos para o Modelo 1 de regressão linear simples. Nos eixos a notação V50_merra e V50_sonda indicam, respectivamente, a velocidade do média do vento à 50 metros para os dados do projeto SONDA e os de reanálises MERRA.....**16**
- Figura 2** – Gráficos para o Modelo 2 de regressão linear. Nos eixos a notação V50_merra e V50_sonda indicam, respectivamente, a velocidade do média do vento à 50 metros para os dados do projeto SONDA e os de reanálises MERRA.....**17**
- Figura 3** – Gráficos para o Modelo 3 de regressão linear. Nos eixos a notação V50_merra e V50_sonda indicam, respectivamente, a velocidade do média do vento à 50 metros para os dados do projeto SONDA e os de reanálises MERRA..... **18**
- Figura 4** – Gráficos para Modelo 4 de regressão linear. Nos eixos a notação V50_merra e V50_sonda indicam, respectivamente, a velocidade do média do vento à 50 metros para os dados do projeto SONDA e os de reanálises MERRA.....**19**

LISTA DE TABELAS

- Tabela 1** – Apresentação dos testes para identificar pressupostos dos modelos de Regressão Linear Simples.....**13**
- Tabela 2** – Modelos de regressão linear obtidos para o intervalo diário (Modelo 1), 5 dias (Modelo 2), 10 dias (Modelo 3), 15 dias (Modelo 4). Para cada modelo são mostrados os coeficientes de determinação, coeficientes de determinação ajustado e o valor-p dos testes para identificação dos pressupostos**15**
- Tabela 3** – Modelos de regressão linear obtidos para o intervalo diário (Modelo 1), 5 dias (Modelo 2), 10 dias (Modelo 3) e 15 dias (Modelo 4). Para cada modelo são mostrados os coeficientes de determinação e de determinação ajustado.....**15**

SUMÁRIO

1 INTRODUÇÃO	11
2 MATERIAL E MÉTODOS	12
2.1 Área de estudo	12
2.2 Análise de regressão	12
3 RESULTADOS E DISCUSSÕES	14
4 CONCLUSÃO	20
REFERÊNCIAS	21

UTILIZAÇÃO DO MÉTODO DE REGRESSÃO LINEAR EM DADOS OBSERVADOS E ESTIMADOS DA VELOCIDADE DO VENTO NA REGIÃO DO SUBMÉDIO RIO SÃO FRANCISCO

USE OF LINEAR REGRESSION METHOD IN OBSERVED AND ESTIMATED WIND SPEED DATA IN THE MIDDLE REGION OF SÃO FRANCISCO RIVER

PallomaKattyAnne de O. S. ^{1*}
Pollyanna Kelly de O. S. ²
Gustavo H. Esteves ^{3*}

RESUMO

O trabalho apresentado trata da estimativa de recursos eólicos em Petrolina-PE, usando a teoria da regressão linear, cujo objetivo é de validar a velocidade do vento através dos dados estimados pelo modelo numérico MERRA (*Modern Era Retrospective Analysis for Research and Application*) com os dados observados da torre anemométrica do projeto SONDA (Sistema de Organização Nacional de Dados Ambientais). São utilizados dados horários da velocidade do vento à 50 metros para o período de 01 de outubro de 2009 a 31 de maio de 2010. A escolha desse período de dados foi motivada pela atuação de intensidade moderada do El Niño (fase positiva do El Niño-Oscilação Sul), que sobre a região do Nordeste do Brasil tende a diminuir os índices de precipitação. Os modelos de regressão linear obtidos para quatro intervalos distintos mostraram que a velocidade do vento observado pela torre anemométrica do projeto SONDA está bem correlacionada com os dados estimados pelo modelo numérico MERRA, principalmente quando a análise é realizada com as médias horárias nos intervalos de 10 e 15 dias, já que a variabilidade dos dados consegue ser suavizada.

Palavras-chave: Projeto SONDA. Dados de Reanálises. El Niño-Oscilação Sul.

ABSTRACT

The present work deals with the estimation of wind resources in Petrolina-PE, using linear regression theory, whose objective was to validate the wind speed through the data estimated by the numerical model MERRA (Modern Era Retrospective Analysis for Research and Application) with the observed data of the anemometric tower of the project SONDA (National System of Environmental Data). Wind speed hourly data are used at 50 meters for the period from October 1, 2009 to May 31, 2010. The choice of this period was motivated by the moderate intensity of El Niño (positive El Niño-Southern Oscillation), that on the region of the Northeast of Brazil tends to decrease the precipitation indices. The linear regression models obtained for four different intervals showed that the wind velocity observed by the anemometric tower of the SONDA project is well correlated with the data estimated by the numerical model

¹Bacharelanda em Estatística do CCT-UEPB. Email: pallomakattyanne7@gmail.com.

² Professora do DE-CCT-UEPB. Email: pollyanna_kelly@yahoo.com.br.

³ Professor do DE-CCT-UEPB. Email: gsteves@uepb.edu.br.

MERRA, especially when the analysis is performed with the hourly means in the intervals of 10 and 15 days, since the variability of the data can be smoothed.

Keywords: SONDA Project. Reanalysis Data. El Niño-Southern Oscillation.

1 INTRODUÇÃO

O Rio São Francisco, também denominado por “Velho Chico”, mostra-se como um dos mais importantes rios do Brasil. A sua nascente começa no estado de Minas Gerais na Serra da Canastra, na localidade do município de São Roque de Minas, percorrendo até o encontro com o oceano, passando por diversas regiões e estados do país.

A cidade de Petrolina, situada no estado de Pernambuco, apresenta dados relevantes que demonstram o progresso econômico advindo do rio São Francisco, maior responsável pela prosperidade das áreas ribeirinhas, estas conhecidas pela denominação de Vale do São Francisco. Atualmente a região é a maior produtora de frutas tropicais do país e vem destacando-se também na área energética devido ao grande número de projetos eólicos que estão sendo desenvolvidos no entorno da região.

Devido à importância econômica da cidade e diante da necessidade de se conhecer o potencial eólico dessa região, que está situada às margens do rio São Francisco, o estudo da velocidade média dos ventos é de grande importância para o melhor aproveitamento desse recurso energético renovável. Segundo Beruski et al. (2009), é a partir desses dados que se torna possível verificar se uma determinada região possui potencial eólico satisfatório, viabilizando tecnicamente a instalação de torres de captação de energia.

A fonte eólica vem ganhando espaço dentro da política de expansão energética brasileira nos últimos anos, atualmente ela representa 8,8% da oferta interna de energia de acordo com o último Balanço Energético Nacional de 2019 ano base 2018 (BEN-2019). O Brasil possui o potencial de vento adequado para exploração eólica e o governo brasileiro tem apoiado o desenvolvimento do setor desde a criação, em 2002, do programa PROINFA (Programa de Incentivo às Fontes Alternativas de Energia Elétrica) com contratação dessa fonte através dos leilões de energia que ocorrem periodicamente, em 2019 houveram dois leilões de Energia Nova A-4 e A-6 (EPE, 2019). Além da capacidade de geração de eletricidade, a energia eólica tem participação fundamental na questão de segurança energética brasileira (Pes, 2015), que com as demais fontes renováveis representam 83,3% da matriz energética do Brasil (BEN-2019).

O uso da teoria de regressão linear para obter estimativas da velocidade média do vento em área de potencial eólico vem sendo empregada por autores como Rosa Filho et. al (2018), Lira et. al (2011) e Camelo et. al (2016), mas ainda são poucos os trabalhos na literatura que utilizam essa técnica para estudar o potencial eólico no nosso país, em especial na região Nordeste.

A partir desta necessidade, o presente trabalho teve o objetivo de avaliar a estimativa da velocidade do vento a partir de dados de reanálises obtidos do modelo MERRA (*Modern Era Retrospective Analysis for Research and Application*) e dados observados da torre anemométrica do projeto SONDA (Sistema de Organização Nacional de Dados Ambientais) do INPE (Instituto Nacional de Pesquisas Espaciais)

através do modelo de regressão linear para dados obtidos na região da cidade de Petrolina, que está situada às margens do Submédio do rio São Francisco.

2 MATERIAL E MÉTODOS

2.1 Área de estudo

O estudo foi realizado na área do município de Petrolina, na margem esquerda do Rio São Francisco, na região do trecho denominado Submédio do São Francisco (CBHSF, 2015). A estação fica próxima à área de divisa com o trecho denominado Médio Do São Francisco (CBHSF, 2015), e a cerca de 60 km do Lago de Sobradinho.

Foram utilizadas duas fontes de dados de vento: a primeira os registros da torre anemométrica situada na cidade de Petrolina, disponibilizados pelo projeto SONDA e a segunda obtida a partir de dados de reanálises do modelo MERRA para a mesma coordenada geográfica. O projeto SONDA foi desenvolvido pelo INPE com objetivo de realizar uma infraestrutura física e de recursos humanos para o melhoramento da base de dados dos recursos de energia solar e eólica no Brasil. Os dados de reanálise MERRA são análises de longo prazo baseadas em uma versão do sistema de assimilação de dados atmosféricos GEOS-5 e foram produzidos em uma grade de $0,5^\circ \times 0,66^\circ$ com 72 camadas.

A torre anemométrica do projeto SONDA está instalada no ponto com as coordenadas geográficas na latitude (graus, minutos, segundos): $09^\circ 04' 08''$ S e longitude (graus, minutos, segundos): $40^\circ 19' 11'' 0$ e sua altitude é 387 metros. Os dados da velocidade e direção do vento e da temperatura apresentam alturas de medição de 25 e 50 metros. Os dados de reanálises do modelo MERRA, possuem as mesmas coordenadas geográficas da estação da torre anemométrica SONDA e foram obtidas as observações da velocidade e direção do vento com alturas de medição de 50 metros.

O período de observação dos dados compreende do dia 1 de outubro de 2009 até 01 de maio de 2010, além disso nesse período houve a ocorrência das fases positiva (El Niño) e moderada do ENOS (El Niño Oscilação Sul) (CLIMATE PREDICTON CENTER, 2019).

2.2 Análise de regressão

Segundo Kazmier (1982) o objetivo principal da análise de regressão linear é prever o valor de uma variável (variável dependente), dado que sejam conhecidos os valores de uma ou mais variáveis associadas (variáveis independentes).

Logo, a análise de regressão estuda a relação entre uma variável dependente e outras variáveis independentes. A relação entre elas é representada por um modelo matemático, que associa a variável dependente com as variáveis independentes. Este modelo é designado por modelo de regressão linear simples (MRLS) se define uma relação linear entre a variável dependente e uma única variável independente. Mas se quer relacionar a variável dependente com mais de uma variável independente, a regressão linear múltipla é utilizada.

O modelo de regressão linear simples pode ser representado por:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, i = 1, 2, \dots, n. \quad (1)$$

em que, n é o tamanho da amostra; Y_i representa o valor da variável dependente, Y , na observação i , $i=1,2,\dots,n$; X_i representa o valor da variável independente, X , na observação i , $i=1,2,\dots,n$; ε_i , $i = 1,2,\dots,n$ são variáveis aleatórias que correspondem ao erro; e por fim β_0 e β_1 são os parâmetros a serem estimados.

De acordo com Charnet et.al (2008) ao se estabelecer esse modelo, dado pela equação (1), pressupõe-se que a relação entre Y e X é linear; os valores de X são fixos (não aleatórios); a média do erro é nula, isto é, $E(\varepsilon_i) = 0$; e para um dado valor de X , a variância do erro ε_i é sempre σ^2 , isto é,

$$Var(\varepsilon_i) = E(\varepsilon_i^2) - [E(\varepsilon_i)]^2 = E(\varepsilon_i^2) = \sigma^2, \quad (2)$$

o que implica em

$$Var(Y_i) = E[Y_i - E(Y_i)]^2 = E(\varepsilon_i^2) = \sigma^2. \quad (3)$$

Diz-se, então, que o erro é homoscedástico, ou que se tem homoscedasticidade (do erro ou da variável dependente). Além disso, o erro de uma observação é independente do erro de outra observação, isto é,

$$Cov(\varepsilon_i, \varepsilon_i') = E(\varepsilon_i \varepsilon_i') - E(\varepsilon_i)E(\varepsilon_i') = E(\varepsilon_i \varepsilon_i') = 0, \text{ para } i \neq i'. \quad (4)$$

Neste tipo de modelagem também se assume que os erros seguem distribuição normal. Logo, combinando as expressões (2), (3) e (4) tem-se $\varepsilon_i \sim N(0, \sigma^2)$ e, portanto, $Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$. A suposição de normalidade é necessária para a elaboração dos testes de hipóteses e obtenção de intervalos de confiança.

Para identificação desses pressupostos associados à modelagem foram utilizados o teste de verificação de normalidade dos resíduos de Shapiro-Wilk (TSW), teste de independência dos resíduos de Durbin-Watson (TDW) e o teste que verifica a igualdade da variância dos resíduos, ou seja, hipótese de homoscedasticidade através do Teste de Bartlett (TB). Na Tabela 1, tem-se a representação resumida dos testes utilizados, sendo que os mesmos serão aplicados ao nível de significância de $\alpha = 0,05$. Além disso, para verificar a qualidade do ajuste do modelo foi utilizado o coeficiente de determinação (MONTGOMERY et. al, 2012).

Tabela 1: Apresentação dos testes para identificar pressupostos dos modelos de Regressão Linear Simples

Teste de Shapiro-Wilk (TSW)	
H ₀ : amostra provém de uma população Normal.	H ₁ : amostra não provém de uma população Normal.
Tomar decisão: se o p-valor for maior do α , ou seja, $p > 0,05$ (não rejeita H ₀).	
Teste Durbin-Watson (TDW)	
H ₀ : os resíduos são independentes	H ₁ : os resíduos não são independentes.
Tomar decisão: se o p-valor for maior do α , ou seja, $p > 0,05$ (não rejeita H ₀).	
Teste Bartlett (TB)	
H ₀ : os resíduos possuem homoscedasticidade.	H ₁ : os resíduos possuem heterocedasticidade.
Tomar decisão: se o p-valor for maior do que α , ou seja, $p > 0,05$ (não rejeita H ₀).	
Fonte: adaptado de Camelo (2016).	

Devido à alta variabilidade dos dados, os modelos de regressão linear serão obtidos para quatro intervalos de tempo: diário, 5 dias, 10 dias e 15 dias. Lira et al. (2011) também utilizou a técnica de suavização para os dados de velocidade média do vento em duas localidades no Ceará obtendo esses quatro intervalos (diário, de 5, 10 e 15 dias).

3 RESULTADOS E DISCUSSÕES

Os dados horários de reanálises do MERRA para a velocidade média do vento à 50 metros foram utilizados para obter o modelo inicial de regressão linear, mas os resultados obtidos nessa escala de tempo não foram satisfatórios (não mostrado). Isso se deve ao fato de que a variável estudada, velocidade média do vento, apresenta máximos e mínimos bem acentuados ao longo do dia e essa alta variabilidade não favoreceu ao modelo com dados horários.

No intuito de suavizar essa variabilidade, calculou-se as médias horárias da velocidade do vento em intervalos diário, de 5, 10 e 15 dias.

Após esse processo, foi obtida a estimação dos parâmetros do modelo de regressão linear simples, pelo método de mínimos quadrados, utilizando os dados provenientes do modelo de reanálises MERRA (variável X) para estimar os dados de velocidade média do vento no mesmo ponto dos dados já existentes do projeto SONDA (variável Y). Por fim, foram obtidos quatro modelos de regressão linear simples utilizando as médias horárias para cada um dos intervalos.

Os modelos de regressão linear simples ajustados para os quatro intervalos são mostrados na Tabela 2. Além dos coeficientes dos parâmetros do modelo também foram obtidos os coeficientes de determinação e o de determinação ajustado. Analisando esses coeficientes podemos perceber que os modelos 2 e 4 são os que apresentam os maiores valores, o Modelo 2 mostra que aproximadamente 87% da variação percentual da velocidade do vento foi explicada pelo MRLS e para o Modelo 4 a variação explicada foi de aproximadamente 91%. Esses coeficientes confirmam a evidência produzida pelos testes t e F que sugere que o modelo de regressão se ajustou bem. Entretanto, para identificação do melhor modelo também é necessário verificar se os pressupostos do MRLS para os resíduos, verificando a sua normalidade, a autocorrelação e a homoscedasticidade não são violados.

As Figuras de 1 a 4 ilustram os gráficos para a dispersão dos dados, para a normalidade, para os resíduos e para os resíduos ajustados. Nos gráficos de dispersão para a velocidade média do vento à 50 metros, elaborado com os dados de reanálise MERRA e com os dados do projeto SONDA, é possível verificar que há indícios de que o grau de associação entre as duas variáveis é positivo, ou seja, à medida que a velocidade do vento obtida pelo modelo estimado MERRA aumenta, a velocidade do vento obtido pelo projeto Sonda também aumenta (Fig. 1-4 a).

Os gráficos para a normalidade dos resíduos mostram que existem pontos que se afastam ligeiramente da diagonal (Fig. 1-4b), mas com apenas a análise gráfica não é possível concluir que os resíduos seguem normalidade. Dessa forma, foi realizado o teste de normalidade de Shapiro-Wilk para os quatro modelos que indicaram que os resíduos parecem ser normalmente distribuídos, já que o valor- p para todos os modelos foi maior do que 0,05 (Tabela 3).

Para checar a autocorrelação dos resíduos foram aplicados o teste de Durbin-Watson nos Modelos de 1 a 4 e os resultados do valor- p mostram que para os Modelos 1 e 2 os resíduos não são independentes, violando assim um dos pressupostos do

MRLS. Entretanto para os Modelos 3 e 4 os resíduos são independentes, já que o valor- p foi maior que 0,05.

Por fim, para analisar o último pressuposto foram analisados se os erros possuem variância constante e se são não correlacionados entre si (Figuras 1-4c.d.) (Tabela 3). Para essa análise foram obtidos os gráficos dos resíduos e dos resíduos ajustados para checar a homoscedasticidade dos resíduos e foi possível perceber que a suposição de homoscedasticidade foi violada para o Modelo 1 e confirmada pelo Teste de Bartlett, implicando na heterocedasticidade dos resíduos com valor- p de aproximadamente 0,001. Para os Modelos de 2 a 4 o pressuposto de homoscedasticidade parece não ter sido violado ao analisar graficamente os resíduos, mas para validar esse resultado o Teste de Bartlett foi aplicado confirmando que a suposição de homoscedasticidade não foi violada.

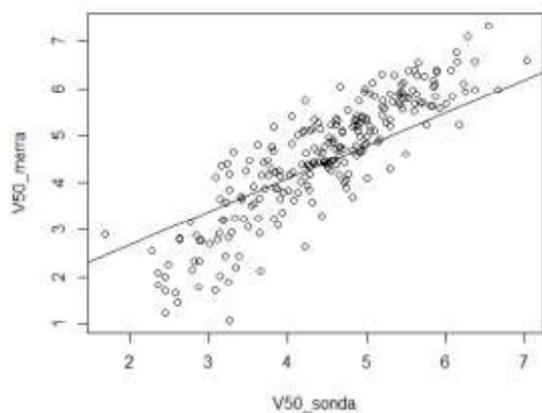
Tabela 2: Modelos de regressão linear obtidos para o intervalo diário (Modelo 1), 5 dias (Modelo 2), 10 dias (Modelo 3) e 15 dias (Modelo 4). Para cada modelo são mostrados os coeficientes de determinação e de determinação ajustado.

Modelo	MRLS	Coefficiente de Determinação	Coefficiente de Determinação Ajustado
Modelo 1	$\hat{Y} = 1,2610 + 0,7024X$	77,02%	77,00%
Modelo 2	$\hat{Y} = 1,1130 + 0,7364X$	87,83%	87,58%
Modelo 3	$\hat{Y} = 1,2157 + 0,7144X$	85,25%	84,58%
Modelo 4	$\hat{Y} = 1,2281 + 0,7094X$	91,14%	90,54%

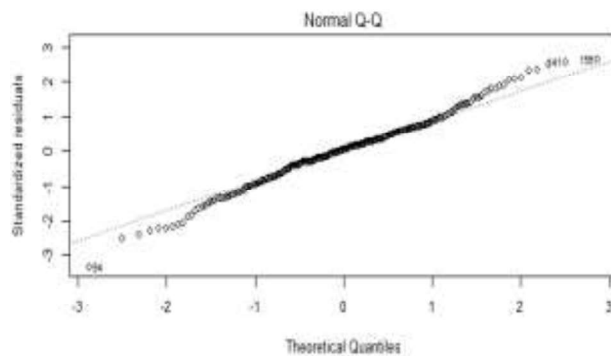
Tabela 3: Modelos de regressão linear obtidos para o intervalo diário (Modelo 1), 5 dias (Modelo 2), 10 dias (Modelo 3) e 15 dias (Modelo 4). Para cada modelo são mostrados os coeficientes de determinação e de determinação ajustado.

Modelo	Teste de Shapiro-Wilk	Teste de Durbin-Watson	Teste de Bartlett
Modelo 1	0,1356	0,00012	0,00130
Modelo 2	0,7262	0,01206	0,0982
Modelo 3	0,2605	0,349	0,2261
Modelo 4	0,5759	0,1159	0,2457

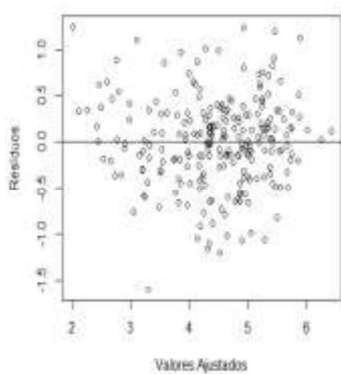
Figura 1: Gráficos para o Modelo 1 de regressão linear simples. Nos eixos a notação V50_merra e V50_sonda indicam, respectivamente, a velocidade do média do vento à 50 metros para os dados do projeto SONDA e os de reanálises MERRA.



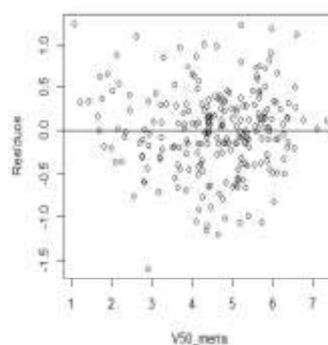
(a) Gráfico de dispersão



(b) Gráfico de normalidade

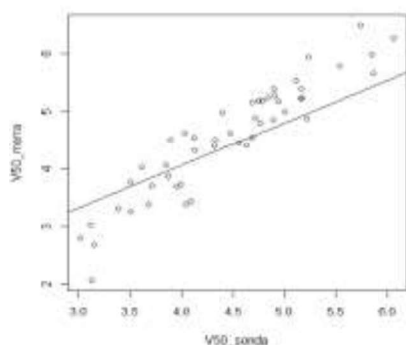


(c) Gráfico para análises dos resíduos

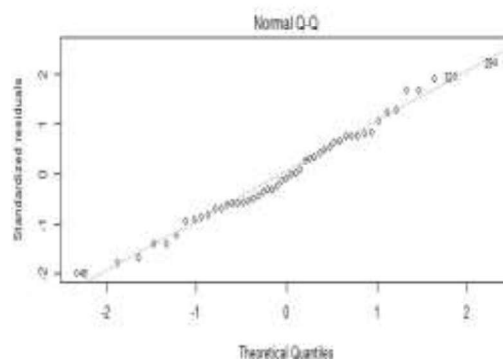


(d) Gráfico para análises dos resíduos

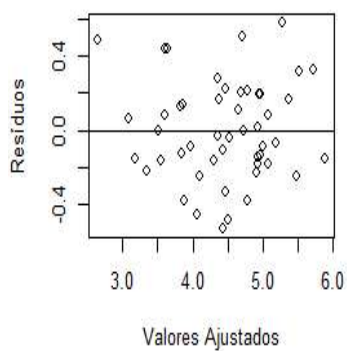
Figura 2: Gráficos para o Modelo 2 de regressão linear. Nos eixos a notação V50_merra e V50_sonda indicam, respectivamente, a velocidade do média do vento à 50 metros para os dados do projeto SONDA e os de reanálises MERRA.



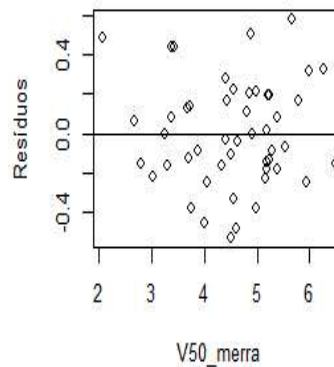
(a) Gráfico de dispersão



(b) Gráfico de normalidade

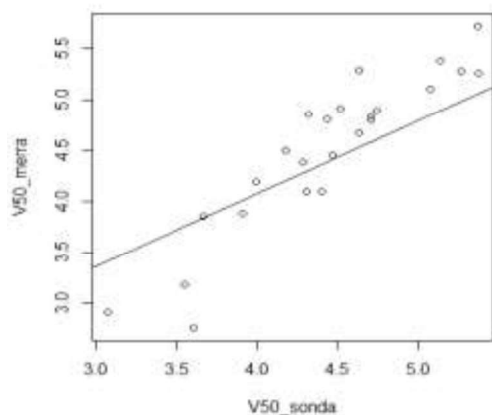


(c) Gráfico para análises dos resíduos

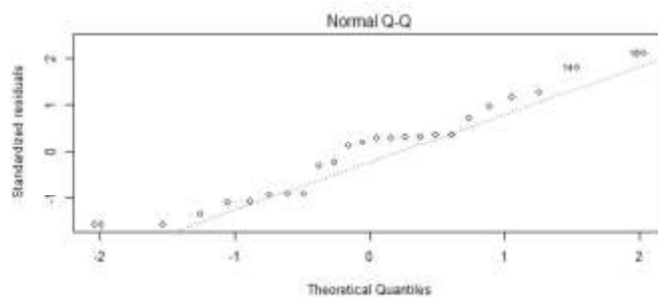


(d) Gráfico para análises dos resíduos

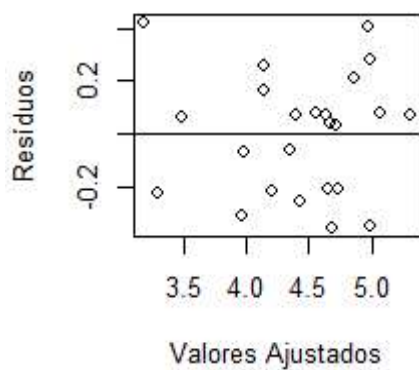
Figura 3: Gráficos para o Modelo 3 de regressão linear. Nos eixos a notação V50_merra e V50_sonda indicam, respectivamente, a velocidade do média do vento à 50 metros para os dados do projeto SONDA e os de reanálises MERRA.



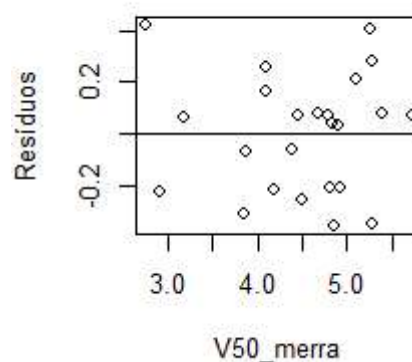
(a) Gráfico de dispersão



(b) Gráfico de normalidade

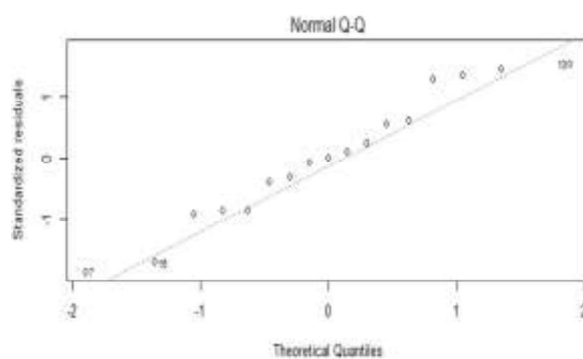
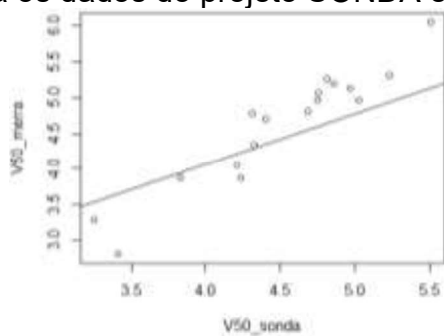


(c) Gráfico para análises dos resíduos



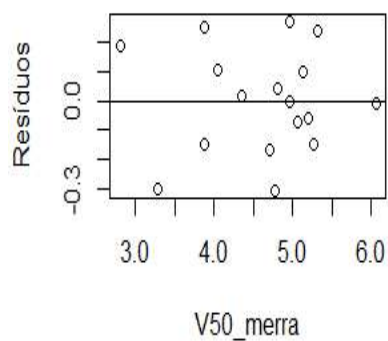
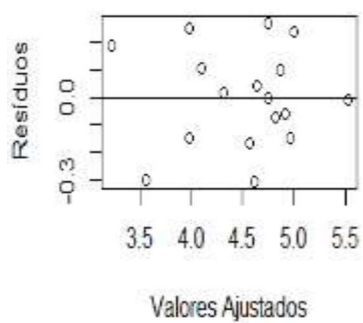
(d) Gráfico para análises dos resíduos

Figura 4: Gráficos para Modelo 4 de regressão linear. Nos eixos a notação V50_merra e V50_sonda indicam, respectivamente, a velocidade do média do vento à 50 metros para os dados do projeto SONDA e os de reanálises MERRA.



(a) Gráfico de dispersão

(b) Gráfico de normalidade



(c) Gráfico para análises dos resíduos

(d) Gráfico para análises dos resíduos

4 CONCLUSÃO

Os modelos de regressão linear apresentados neste estudo, mostraram desempenho satisfatórios entre os dados de velocidade do vento observados pelo projeto SONDA e os estimados numericamente pelo modelo MERRA. Houve um bom índice de concordância entre as séries de dados estimados e a série de dados observados e seus respectivos coeficientes de correlação, principalmente entre as médias horárias da velocidade do vento à 50 metros de 10 e 15 dias. Ocasionalmente, principalmente pela suavização da variabilidade da velocidade do vento, impactando assim em baixos valores dos erros quantificados. Além disso, em ambos os casos (com médias tomadas de 10 e 15 dias) os pressupostos do modelo de regressão linear simples não foram violados. Por outro lado, com a média de velocidade do vento à 50 metros tomada de modo diário e com 5 dias, os modelos de regressão linear simples apresentaram violação nos pressupostos, o que mostra a inviabilidade deste tipo de abordagem nestes casos.

Dessa forma, conclui-se que a qualidade do ajuste do modelo melhora à medida que o tempo entre as médias horárias aumenta. No caso deste trabalho, as médias tomadas em 10 e 15 dias mostraram bons ajustes. Tais resultados apontam que a estimativa dos recursos eólicos para uma dada região, neste caso Petrolina-PE, pode ser obtida utilizando dados de modelagem numérica do tipo MERRA na ausência de dados observados no local do sítio eólico, o que pode ser muito útil em casos onde não seja possível se tomar tais observações para avaliação de potencial eólico.

REFERÊNCIAS

BERUSKI, G. C.; LEITE, M. L.; VIRGENS FILHO, J. S.; ADACHESKI, P. A.; OLIVEIRA, A. V. Análise probabilística da velocidade média e caracterização da direção predominante do vento no município da Lapa/PR. R. RA´EGA, n. 17, p. 65-75, 2009.

CAMELO, H.N.; LUCIO, P.S.; GOMES, O.M.; LEAL JUNIOR, J.B.V. Predição de velocidade do vento em municípios do Nordeste brasileiro através de regressão linear e não linear para fins de geração eólica. Revista Brasileira de Geografia Física, v.9, n.03, p.927-939, 2016.

CBHSF – COMITÊ DA BACIA HIDROGRÁFICA DO RIO SÃO FRANCISCO. Plano de recursos hídricos da bacia hidrográfica do rio São Francisco 2016-2025. v. 1, 2015. Disponível em: <http://cbhsaofrancisco.org.br/planoderecursoshidricos/wp-content/uploads/2015/04/RP2_V1_Diagnostico_Consolidado.pdf>. Acesso em: 14 abr. 2018.

CHARNET, R.; FREIRE, C.A.L.; CHARNET, E.M.R.; BONVINO, H. Análise de Modelos de Regressão Linear: com aplicações. 2 ed. Campinas: Editora da Unicamp, 2008.

CLIMATE PREDICTION CENTER. In: National Water Service. Disponível em: <https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php> Acesso em 08 out. 2019.

EPE – EMPRESA DE PESQUISA ENERGÉTICA. Balanço Energético Nacional. 2019. Disponível em: < <https://ben.epe.gov.br/> >. Acesso em: 08 dez.2019.

EPE – EMPRESA DE PESQUISA ENERGÉTICA. Leilões de Energia. 2019. Disponível em: < <http://www.epe.gov.br/pt/leiloes-de-energia/leiloes>>. Acesso em: 08 dez.2019.

KAZMIER, Leonard J. Estatística aplicada à economia e a administração. 1 ed. São Paulo: McGraw-Hill do Brasil, 1982.

LIRA, M.A.T.; SILVA, E.M.; ALVES, J.M.B. Estimativa dos Recursos Eólicos no Litoral Cearense Utilizando a Teoria da Regressão Linear. Revista Brasileira de Meteorologia, v.26, n.3, p.349-366, 2011.

MONTGOMERY, D.C.; PECK, E.; VINING, G. G. *Introduction to Linear Regression Analysis*. 5 ed. New Jersey: John Wiley and Sons, 2012.

PES, M. P. 2015. Impactos das variações e mudanças climáticas sobre os ventos extremos e seus efeitos no setor elétrico brasileiro. Tese de Doutorado do Curso de Pós-Graduação em Ciência do Sistema Terrestre, Instituto Nacional de Pesquisas Espaciais – INPE, São José dos Campos, São Paulo, Brasil. 174p.

ROSA FILHO, C.D.; PARISI, S.A.; SOARES, W.A. Avaliação da distribuição estatística e elaboração de modelo de regressão múltipla linear da velocidade média do vento. *Journal of Environmental Analysis and Progress*, v.03, n.01, p.24-36, 2018.

AGRADECIMENTOS

À Deus, por sempre me proporcionar trilhar nos melhores caminhos.

À meu orientador, Dr. Gustavo Henrique Esteves, por me proporcionar os ensinamentos no caminhar da graduação e por aceitar minha orientação.

À professora e irmã, Dra. Pollyanna Kelly de Oliveira Silva, por me apoiar e incentivar durante todos os períodos do curso. Além de ser coorientadora desse trabalho, cedendo o banco de dados e orientando sobre o tema desse estudo.

Aos professores do Departamento de Estatística, fundamentais na construção do aprendizado.

Aos membros da banca examinadora por aceitarem participar do processo de avaliação e aprimoramento desse trabalho.

Enfim, a todos que contribuíram para a realização deste trabalho.