

Diogo Friggo

Estimativa da produção de energia de um parque eólico por meio de modelo estocástico

Brasil

2019

Diogo Friggo

Estimativa da produção de energia de um parque eólico por meio de modelo estocástico

O presente trabalho propõe um modelo que represente a estrutura de séries temporais de velocidade de vento de modo a fazer previsões sobre seu comportamento futuro tendo como objetivo estimar a energia que seria gerada por uma turbina eólica em um dado período.

Universidade Federal do Rio Grande do Sul – UFRGS

Trabalho de Diplomação em Engenharia Física

Orientador: Carlo Requião

Brasil

2019

Diogo Friggo

Estimativa da produção de energia de um parque eólico por meio de modelo estocástico/ Diogo Friggo. – Brasil, 2019-

58 p. : il. (algumas color.) ; 30 cm.

Orientador: Carlo Requião

Graduação – Universidade Federal do Rio Grande do Sul – UFRGS
Trabalho de Diplomação em Engenharia Física, 2019.

1. Energia Eólica. 2. Autoregressive Integrated Moving Average (ARIMA). 3. Generalized Autoregressive Conditional Heteroscedasticity (GARCH). I. Universidade Federal do Rio Grande do Sul. II. Engenharia Física.

Diogo Friggo

Estimativa da produção de energia de um parque eólico por meio de modelo estocástico

O presente trabalho propõe um modelo que represente a estrutura de séries temporais de velocidade de vento de modo a fazer previsões sobre seu comportamento futuro tendo como objetivo estimar a energia que seria gerada por uma turbina eólica em um dado período.

Brasil
2019

Resumo

Todos os agentes envolvidos no planejamento de um parque eólico precisam garantir, com um certo nível de certeza, quanta energia este poderá gerar. Investidores precisam avaliar o risco associado ao financiamento da construção do parque em relação a um retorno futuro. Produtores precisam garantir um determinado nível de produção de energia com base em acordos mensais com compradores. Operadores de subestações, por outro lado, precisam ter uma estimativa da produção em resolução horária de modo a evitar perdas elétricas e para que possam atender adequadamente a demanda por energia a qual sabe-se que varia ao longo do dia de maneira previsível e conhecida a partir de dados históricos de consumo por uma dada região.

Atualmente existem metodologias bem estabelecidas na indústria para a estimativa do recurso eólico de longo prazo e da geração de energia de longo prazo, geralmente para um horizonte de 20 anos. Elas se baseiam no fato de que, a longo prazo, a distribuição de probabilidade da velocidade do vento aproxima-se de uma distribuição de Weibull ([MENČÍK, 2016](#)) a qual é então convertida em uma distribuição de probabilidade de energia por meio de uma curva de potência de uma determinada turbina. Essa distribuição de densidade de probabilidade é a mais difundida na análise de dados de vento ([P., 2003](#)), tendo sido utilizada para estimar tanto o recurso eólico ([AN., 2003](#)) quanto a produção de energia ([M HOOSHMAND R-A; A-H, 2015](#)).

O sucesso das metodologias supracitadas é explicado pelo fato de que um longo período de dados é utilizado para produzir uma previsão de extensão temporal similar. Isso tem o efeito de remover flutuações que são observados entre diferentes períodos. Com tantos dados disponíveis, uma distribuição pode ser adequadamente ajustada, a qual fidedignamente representa a verdadeira distribuição de velocidade do vento. Há uma vasta quantidade de trabalhos comerciais que confirmam essa hipótese. No entanto, há uma crescente necessidade de obter-se previsões cada vez mais próximas do presente. Sempre que a produção de energia prevista não corresponde a energia efetivamente gerada, todas as partes envolvidas perdem dinheiro. Mesmo que mais energia do que o esperado seja gerada, ainda assim é comum que a subestação que receberá essa energia não esteja preparada para receber energia em excesso.

Um passo inicial na tentativa de encontrar um equilíbrio na cadeia de produção e distribuição de energia eólica consiste em primeiramente prever com um certo grau de confiança a velocidade do vento futura, dado que essa é a variável que é a maior responsável pela incerteza na produção de energia. Este trabalho propõe o uso de modelos autoregressivos,

em particular modelos GARCH, de modo a produzir previsões confiáveis em um horizonte de horas até meses.

Palavras-chave: energia eólica, previsão de séries temporais, modelos autoregressivos

Abstract

Every stakeholder of a windfarm must ensure, within a specified confidence level, how much energy that windfarm may generate. Investors must balance the risk that goes along with financing such a project and the expected return of investment. Developers are held accountable to a certain level of energy production that is previously agreed upon, usually on a monthly basis, but also yearly. Grid operators on the other hand demand an hourly estimate of production so that they can match offer to demand with minimal losses due to circuit inadequacy. Energy demand follows a quite predictable pattern that may be easily anticipated based on past data. That is much less so the case for energy offer which, besides coming from a variety of sources, also brings along a lot of complexities, some of which are discussed on this text for the case of energy derived from wind.

Nowadays there are well-established methodologies in the field of wind energy forecasting. Usually the time horizon involved is of 20 years. These methodologies rest on the assumption that a long-tailed Weibull distribution ([MENČÍK, 2016](#)) may always be fit to long-term wind speed. This distribution is considered representative of 20 years into the future and is then converted into energy ([M HOOSHMAND R-A; A-H, 2015](#)) by mapping it through a power curve specific to a manufacturer's turbine model. Its widespread use is justified by the vast amount of commercial work done on the hypothesis that a Weibull distribution best describes long-term wind speed ([AN., 2003](#)).

The success of the aforementioned methodologies may be explained by the fact that a very long period of data is used to achieve a similarly long forecast. This has the effect of averaging out fluctuations observed from one period to the next. With so much data available a distribution fit may be found which reliably represents the true wind speed distribution. There are a plethora of commercial work that support this view. However, there's a growing need to bring the forecast horizon ever closer to the present. Whenever the forecast energy production doesn't match the realized one, all stakeholders lose money. Even if more energy than expected is produced there's money lost because the target substation may not be able to handle the excess energy.

An initial step into finding the optimal balance in this wind energy chain is to first confidently predict future wind speed since this is the variable that is mostly responsible for the uncertainty in energy production. This work pursues the use of autoregressive models, particularly of GARCH models, in order to achieve reliable forecasts with a time horizon of the order of hours to months.

Keywords: wind energy, time-series forecasting, autoregressive models

Listas de ilustrações

Figura 1 – Comparação entre a composição da matriz energética brasileira no ano de 2017 (esquerda) e no ano de 2018 (direita)	12
Figura 2 – Ranking de capacidade eólica instalada acumulada	13
Figura 3 – Evolução da capacidade instalada de parques eólicos no Brasil	13
Figura 4 – Curva de potência de uma turbina genérica. Fonte: Wikimedia Commons	15
Figura 5 – À esquerda velocidade do vento em m/s para uma região no sul do Ceará e à direita o preço de fechamento das ações da google para 120 dias de 2013. Fonte: autoria própria, dados (ECMWF, 2017).	16
Figura 6 – Chapada	18
Figura 7 – Variação da elevação na direção preferencial de escoamento do vento. .	19
Figura 8 – Alguns nós da série ERA 5 ao sul do Ceará, Brasil. Google earth V 7.3.2.5776. (14 de Dezembro, 2015). Rio Grande do Sul, Brasil. $7^{\circ} 14' 11,54''$ S, $40^{\circ} 04' 12,31''$ W, Eye alt 1808,45 km.	22
Figura 9 – Velocidade do vento registrada por satélite na região de interesse nos anos de 2017 e 2018	23
Figura 10 – Histograma de velocidades do nó noroeste da série de dados modelo. .	24
Figura 11 – Gráfico de Cullen e Frey para os dados do nó noroeste da série de dados modelo.	25
Figura 12 – Rosa dos ventos	26
Figura 13 – Chapada	26
Figura 14 – Chapada	27
Figura 15 – O método de Box-Jenkins.	32
Figura 16 – Últimos dois anos da série de vento	33
Figura 17 – Sazonalidade removida	34
Figura 18 – Sazonalidade removida por meio de uma transformação de Box-Cox .	35
Figura 19 – Gráficos da Função de Autocorrelação (ACF) e Função de Autocorrelação Parcial (PACF)	37
Figura 20 – Gráficos da Função de Autocorrelação (ACF) e Função de Autocorrelação Parcial (PACF)	38
Figura 21 – Modelo LSTM aplicado a série. Velocidade no eixo y. Tempo em horas no eixo x.	39
Figura 22 – 3 últimas semanas de dados da série. Antes da diferenciação (superior) e após (inferior).	40
Figura 23 – Gráficos da Função de Autocorrelação e Autocorrelação Parical para 2 semanas de dados da série.	41
Figura 24 – Chapada	43

Figura 25 – Condições de regime estacionário e invertibilidade para um modelo ARIMA(6,1,0).	45
Figura 26 – Condições de regime estacionário e invertibilidade para um modelo ARIMA(6,1,0).	46
Figura 27 – Condições de regime estacionário e invertibilidade para um modelo ARIMA(6,1,0).	47
Figura 28 – Evolução dos parâmetros p e q de um modelo ARIMA ao longo do tempo.	49
Figura 29 – Evolução dos parâmetros p e q de um modelo ARIMA ao longo do tempo.	50
Figura 30 – Evolução dos parâmetros p e q de um modelo ARIMA ao longo do tempo.	51
Figura 31 – GARCH	52
Figura 32 – GARCH	53
Figura 33 – Curva de potência de um aerogerador genérico.	54
Figura 34 – Curva de potência de um aerogerador genérico.	55
Figura 35 – Regiões correspondentes a dados de velocidade do vento do banco de dados ERA5 da ECWMF. Mapa disponível em TDEF Map	56

Lista de abreviaturas e siglas

ECMWF European Center for Medium-Range Weather Forecast

ARMA Autoregressive Moving-Average

Sumário

Introdução	12
I ESTUDO DE CASO	17
1 CARACTERIZAÇÃO DA REGIÃO	18
2 A SÉRIE TEMPORAL MODELO	21
3 CARACTERIZAÇÃO DOS DADOS	23
3.1 Distribuição de probabilidades	23
3.2 Sazonalidade	25
3.3 Horizonte de previsão	27
3.4 Métodos	28
II IMPLEMENTAÇÃO	29
4 BOX-JENKINS	31
4.1 Análise visual	33
4.2 Regime estacionário	34
4.3 Autocorrelação	36
4.4 Redes neurais recorrentes	37
4.5 Janela de dados	39
4.6 Medidas de qualidade de previsões	41
4.6.1 Erro médio absoluto (MAE)	42
4.6.2 Raíz do quadrado da média do erro (RMSE)	42
4.6.3 Erro absoluto médio escalonado (MASE)	42
4.6.4 Validação cruzada	43
4.7 Estabilidade	44
4.8 Previsão	45
5 MODELO AUTOREGRESSIVO VARIÁVEL	49
6 GARCH	51
6.1 Base horária	51
6.2 Base mensal	52

7	CONVERSÃO EM ENERGIA	54
8	VALIDAÇÃO COM OUTRAS SÉRIES	56
	REFERÊNCIAS	58

Introdução

Ao contrário de fontes de energia tradicionais tais como petróleo ou carvão, o vento é uma fonte de energia limpa e renovável. A energia eólica é, em princípio, inesgotável, pois é função apenas da energia cinética das moléculas presentes na atmosfera, as quais, em última instância, estão em movimento, principalmente, devido a radiação solar e atração gravitacional entre corpos massivos que rege a dinâmica de movimento da Terra e consequentemente sua atmosfera. Ser inesgotável não significa que seja infinita. Existe um limite de quanta energia se pode extrair da atmosfera, no entanto, por ser renovável, essa energia é constante e naturalmente reabastecida.

Diante da finitude das principais fontes de energia em uso no mundo, a geração de energia eólica, assim como de energia solar, tem crescido continuamente ao redor do mundo. Na imagem abaixo é possível constatar a tendência de diminuição do uso de fontes de energia não renováveis e aumento das renováveis, principalmente eólica e solar. Em 2017, na imagem a esquerda, a energia eólica era responsável por 8,1% da matriz energética do país. Em 2018 esse número subiu para 9% enquanto que quase todas as outras fontes recuaram.

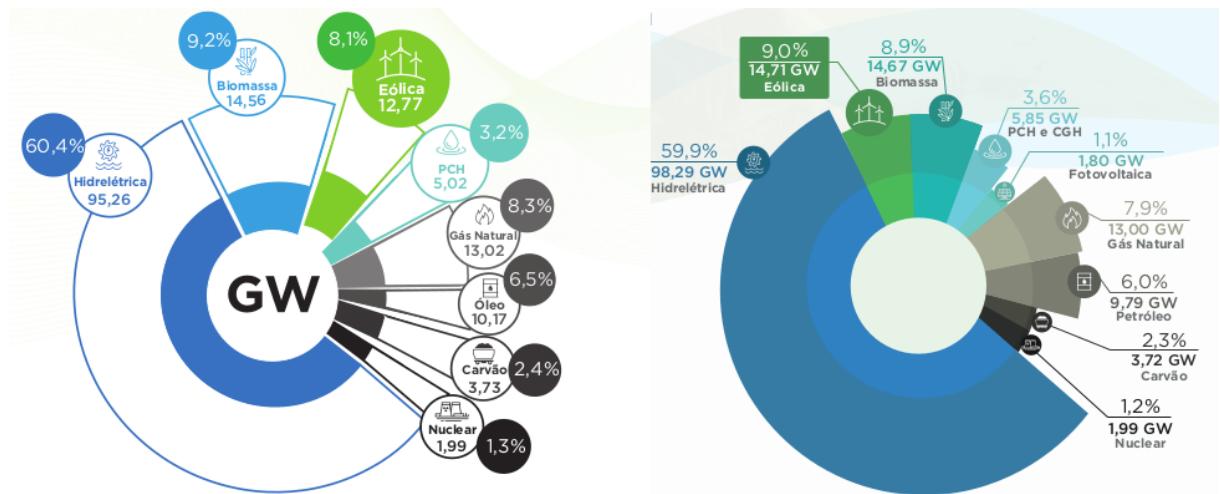


Figura 1 – Comparação entre a composição da matriz energética brasileira no ano de 2017 (esquerda) e no ano de 2018 (direita)

O Brasil conta com uma capacidade instalada de energia eólica de 12,76 GW em seus mais de 500 parques eólicos e figura mundialmente como 8º maior produtor, superando países desenvolvidos como o Canadá.

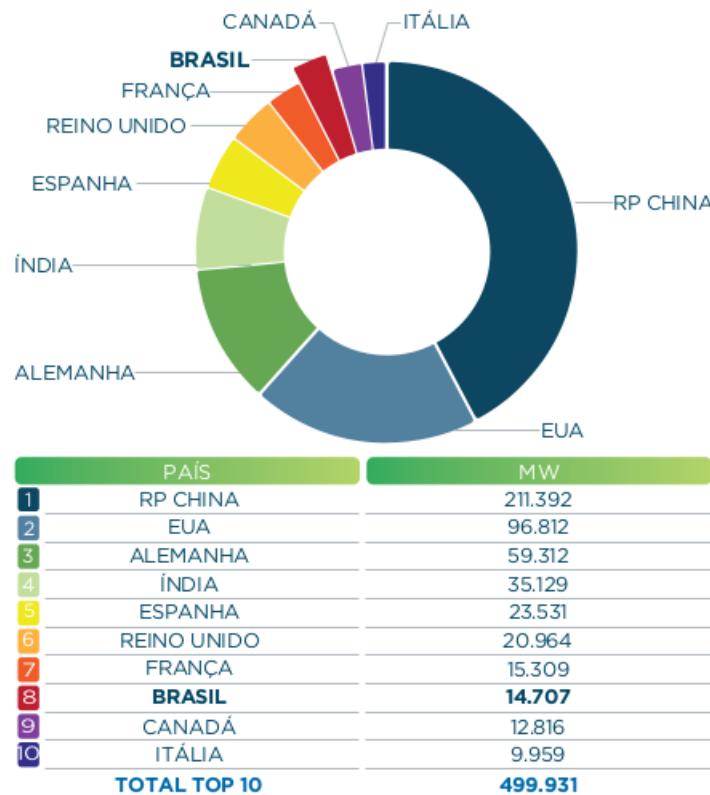


Figura 2 – Ranking de capacidade eólica instalada acumulada

O crescimento da captação de energia eólica no país vem crescendo constatadamente, adicionando em média 2GW de capacidade instalada (máxima capacidade de produção) a cada ano desde 2013. Hoje essa fonte de energia é responsável, em média, por 11% do abastecimento do país e 60% do Nordeste.



Figura 3 – Evolução da capacidade instalada de parques eólicos no Brasil

A principal fonte de energia elétrica do país vêm de hidrelétricas. Em períodos de estiagem, essa fonte é complementada por termoelétricas, as quais usam combustíveis fósseis para gerar energia. O seu alto custo tem impacto direto no consumidor final que

paga mais pela mesma energia.

A variabilidade do recurso eólico representa um dos maiores desafios ao seu aproveitamento. A previsão de que a maré será alta ou baixa é aceita como verdade sem mais questionamentos, enquanto que a previsão de que, devido ao movimento de massas de ar, provavelmente irá chover é encarada com forte suspeita, pois, ao contrário de outras fontes naturais de energia, tais como a energia hídrica ou solar, a energia eólica apresenta grande variabilidade (ACKERMANN, 2012), sendo, portanto, de maior dificuldade a sua previsão. Dessa forma, é de grande interesse que se entenda o seu comportamento e se desenvolva métodos para sua previsão.

Os principais fatores que influem na geração de energia eólica são a densidade do ar e a velocidade do vento. A potência (em watts) gerada por uma turbina é dada pela equação (CEPEL, 2001) a qual é fornecida por fabricantes de turbinas eólicas:

$$P = \frac{1}{2} \rho \frac{\pi D^2}{4} \nu^3 C_p \eta \quad (1)$$

P = potência elétrica na altura do cubo rotor [W]

$$\rho = \text{densidade do ar} \left[\frac{kg}{m^3} \right]$$

D = diâmetro do rotor [m]

$$\nu = \text{velocidade do vento} \left[\frac{m}{s} \right]$$

C_p = coeficiente aerodinâmico de potência do rotor [W]

η = eficiência do conjunto gerador/transmissão

O fato de a potência variar com o cubo da velocidade do vento evidencia a importância da determinação precisa dessa grandeza.

A conversão entre velocidade e potência é feita por uma curva de potência fornecida pelo fabricante da turbina utilizada. Ela representa a potência que seria gerada para cada velocidade incidente perpendicularmente às pás da turbina. Uma vez feita a previsão da velocidade do vento utilizar-se-a, neste trabalho, uma curva de potência genérica como a exibida abaixo para obter a previsão em energia ao invés de velocidade do vento. Observa-se que baixas velocidades de vento resultam em baixa produção de energia e que a velocidades muito altas a produção é constante. Para velocidades muito intensas a turbina é temporariamente desligada para evitar danos estruturais.

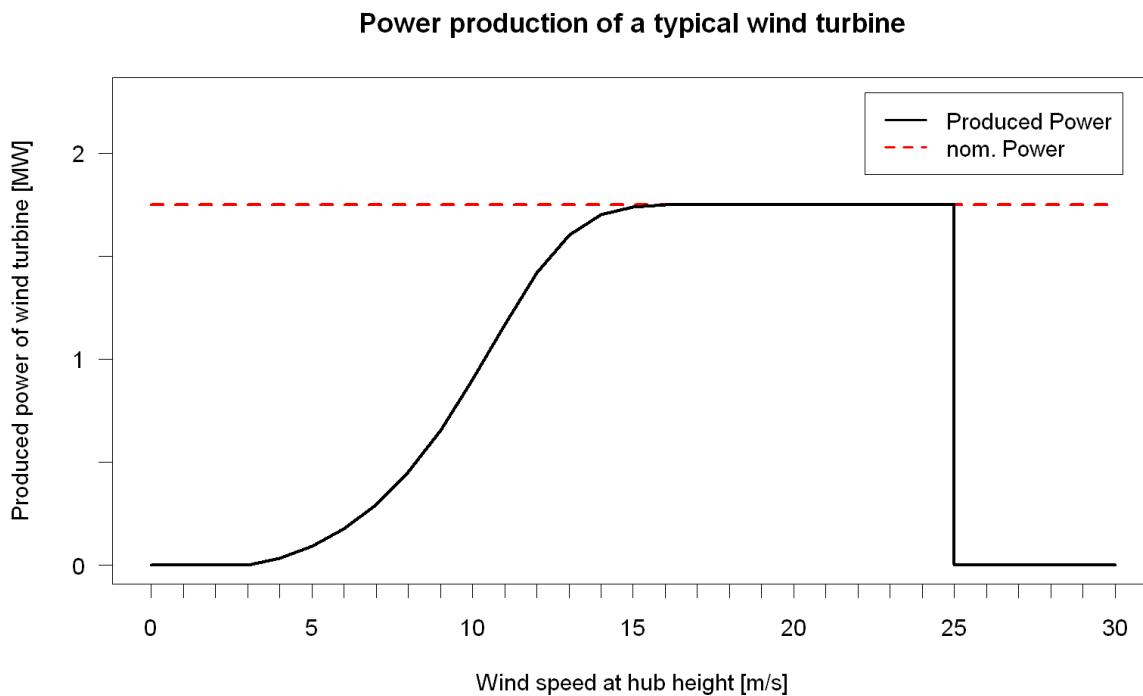


Figura 4 – Curva de potência de uma turbina genérica. Fonte: [Wikimedia Commons](#)

Na equação 1, exceto pela velocidade do vento, os demais parâmetros são controlados ou apresentam pouca variabilidade como a densidade do ar, por exemplo, que é relativamente constante para um dado macroclima. A velocidade do vento, por outro lado, tem caráter fortemente estocástico, ou seja, seu comportamento está associado a um processo aleatório. A sua previsão demanda o uso de modelos matemáticos que levem em conta tal comportamento.

Fenômenos estocásticos são comuns na natureza: o movimento errático que uma partícula macroscópica sofre ao ser imersa num fluído (composto por partículas microscópicas) conhecido como movimento Browniano; o decaimento radiativo de átomos em que não se sabe em qual momento dado átomo emitirá radiação (conhece-se apenas uma taxa característica de emissão).

Exemplos associados a atividade humana também são comuns sendo a evolução temporal do valor de ativos econômicos o exemplo mais marcante e para cujo entendimento se desenvolveu um arcabouço matemático sofisticado. Esse conhecimento pode ser utilizado para estudar quaisquer outros fenômenos estocásticos. A imagem traz dois exemplos de séries estocásticas: uma série de dados de vento (esquerda) e uma série de dados financeira (direita). Os métodos desenvolvidos para estudar séries financeiras podem ser usados com sucesso para séries de velocidade do vento.

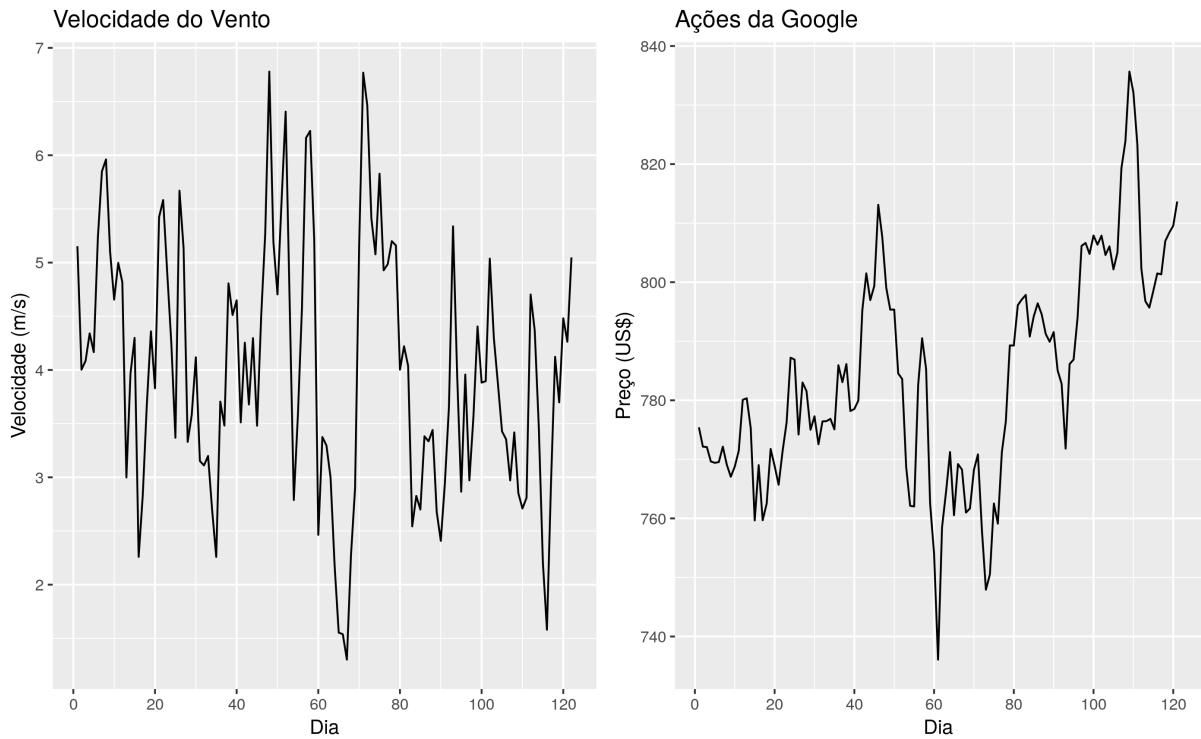


Figura 5 – À esquerda velocidade do vento em m/s para uma região no sul do Ceará e à direita o preço de fechamento das ações da google para 120 dias de 2013. Fonte: autoria própria, dados ([ECMWF, 2017](#)).

Yule e Walker representaram uma grandeza estocástica como resultado de choques aleatórios descritos por um filtro linear aplicado em ruído branco:

$$v_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots$$

Essa ideia dá origem a modelos autoregressivos que descrevem a estrutura de séries temporais levando em conta termos anteriores da própria série. Box e Jenkins descrevem esse método como "deixar que os dados falem por eles próprios", ou seja, nenhuma variável exógena é empregada, apenas a própria série. Esse método tem sido empregado com sucesso para prever o comportamento futuro de séries temporais em diversas áreas. Diversas variações do modelo original foram desenvolvidas ao longo dos anos. O conjunto de todas essas variações são denominadas modelos ARIMA, um acrônimo do inglês *autogressive integrated moving average*, que significa modelo autogressivo (AR) integrado (I) de média móvel (MA).

Parte I

Estudo de caso

1 Caracterização da região

De modo a desenvolver as ideias desse trabalho escolheu-se arbitrariamente uma região localizada no estado do Ceará. Optou-se pela região nordeste por ser a região onde há mais investimentos em energia eólica e que responde pela maioria dos parques em operação no país. Os métodos desenvolvidos são, no entanto, aplicáveis a qualquer região, pois têm como entrada apenas uma série temporal.

A região de interesse encontra-se numa chapada elevada, um local com topografia muito favorável a instalação de um parque eólico sabe-se da dinâmica de fluídos que o escoamento é acelerado quando exposto a um diferencial de elevação. Na imagem abaixo pode-se perceber o quanto elevada a chapada é em relação ao seu redor.

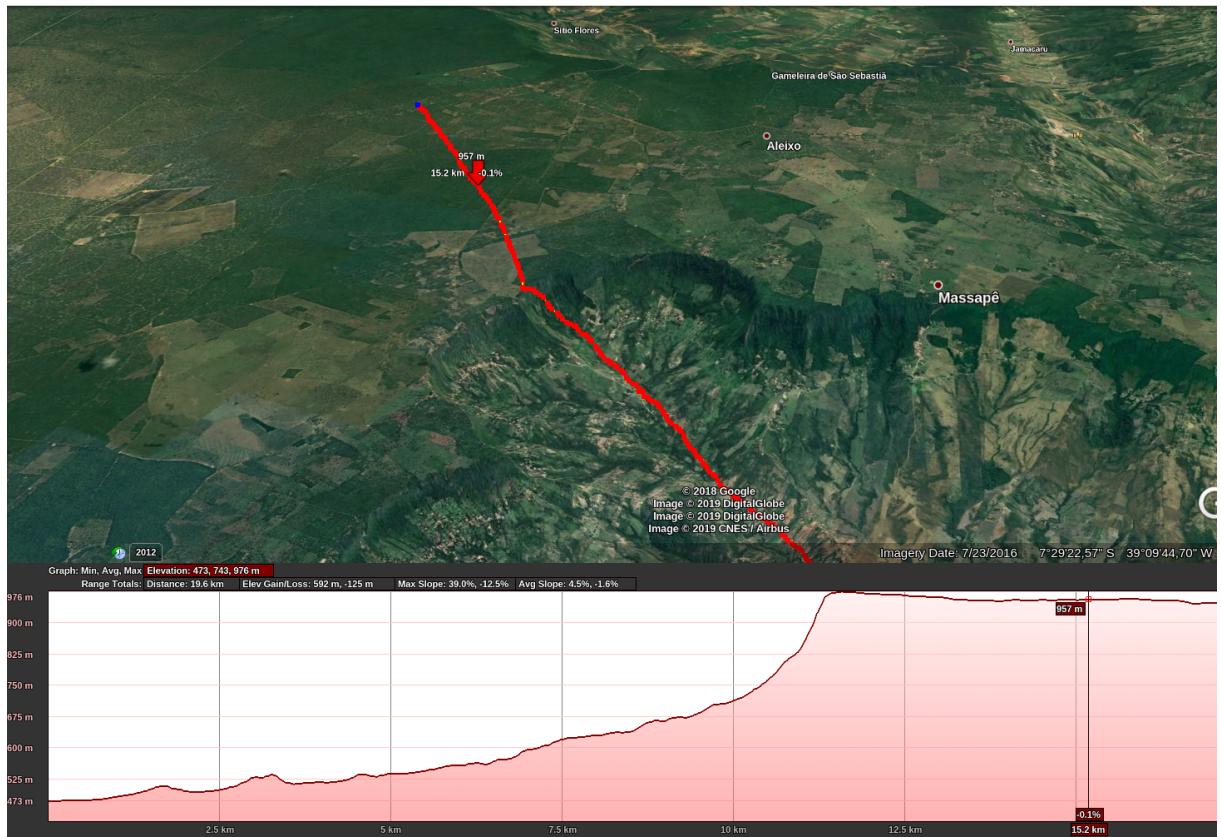


Figura 6 – Chapada

Detalhes da topografia da região pode ser visualizados na imagem abaixo a qual exibe em detalhes a elevação por meio de linhas de contorno além dos declives e a rosa dos ventos. Uma região plana facilita as simulações computacionais necessárias à extração do recurso eólico nas posições de torres de medição às posições de cada turbina.

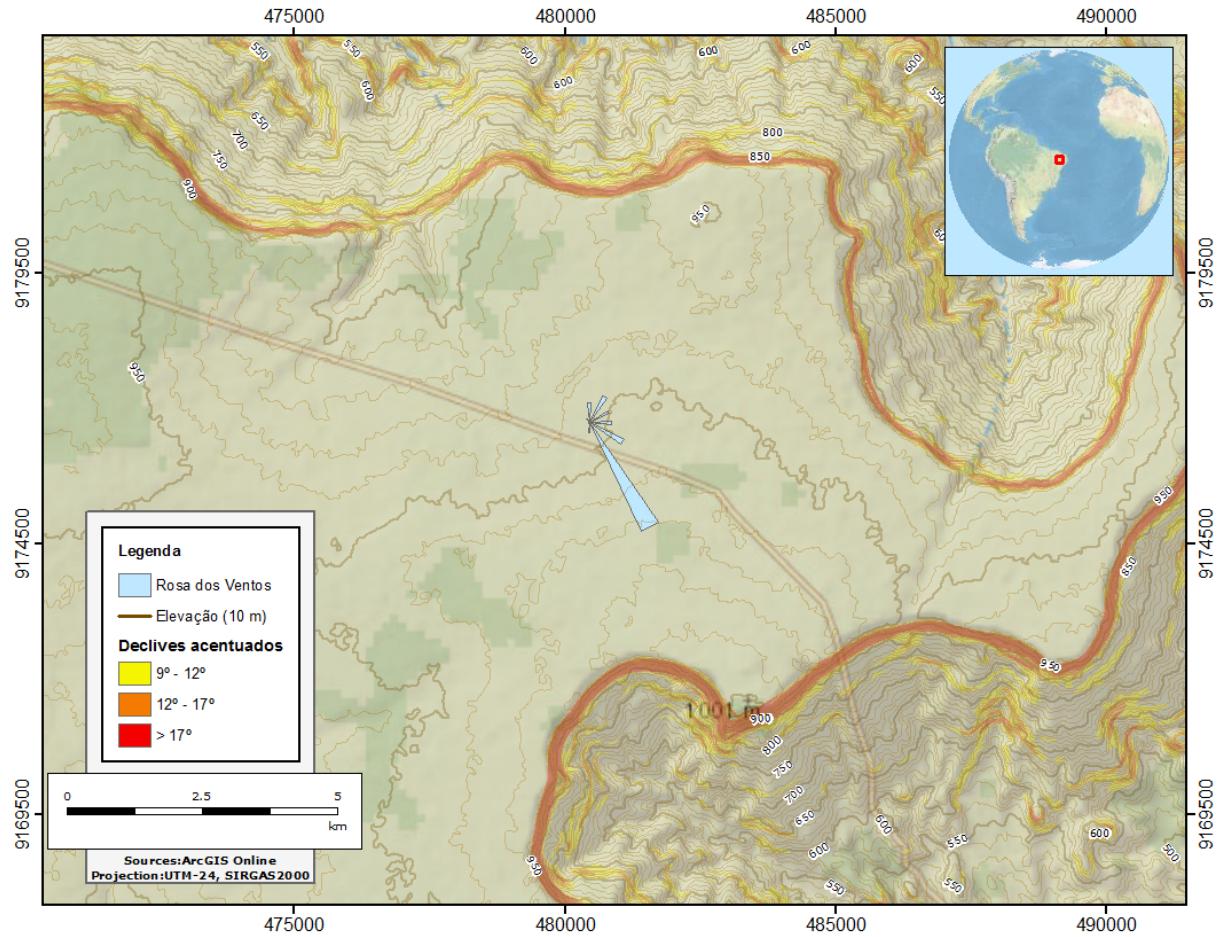


Figura 7 – Variação da elevação na direção preferencial de escoamento do vento.

Na camada da atmosfera próxima ao solo, a velocidade do vento aumenta com a altura conforme uma lei de potência empírica:

$$\frac{v}{v_r} = \left(\frac{h}{h_r} \right)^\alpha \quad (1.1)$$

v = velocidade do vento à altura h

v_r = velocidade do vento à altura de referência h_r

α = coeficiente de cisalhamento, parâmetro estimado a partir dos dados

Em suma, a partir da imagem acima, é possível determinar as melhores posições para se colocar turbinas de um parque eólico:

- Na chapada, onde a velocidade do vento é maior devido a sua elevação de acordo com a lei de potência

- Nas zonas de baixa declividade para facilitar as simulações computacionais
- Nas bordas devido ao diferencial de elevação que causa aceleração do vento
- Orientadas a aproximadamente 150° em relação do norte verdadeiro de modo a captar o vendo oriundo de sua direção preferencial

2 A série temporal modelo

Para exemplificar a elaboração do modelo proposto, utilizou-se uma série de dados disponibilizada publicamente pela organização européia ECWMF. Essa série conta com uma vasta gama de grandezas físicas medidas por satélite oriundas de diversas fontes. Esses dados são agregados e tratados de modo a melhorar sua qualidade. Existem diversas outras séries de dados de satélite com o mesmo propósito, tal como a série ERA-Interim (antecessora da ERA 5) ou as séries MERRA e MERRA 2 produzidas pela NASA. No entanto, observa-se que a qualidade da fonte ERA 5 é muito superior. Isso é empiricamente constatado devido a excelente correlação com dados de torres de medição instaladas no solo. Essa melhoria deve-se tanto a melhorias técnicas na assimilação e agregação de dados quanto a maior resolução espacial da série ERA 5: enquanto MERRA e MERRA 2 apresentam uma resolução espacial de 50 km em latitude e longitude, ERA 5 apresenta 30 km.

A grandeza de interesse para esse trabalho é a velocidade do vento a 100 m de altura em relação ao solo. Essa é a altura disponibilizada pela fonte ERA 5. Outras alturas estão disponíveis mas são obtidas por interpolação e portanto são de qualidade inferior. Essa é uma altura adequada pois a altura do cubo rotor de aerogeradores comerciais varia tipicamente entre 80 m e 130 m.

A série ERA 5, assim como as outras mencionadas, possuem dados para todo o globo, dividindo-o em uma rede de nós com resolução de 30 km x 30 km em latitude e longitude respectivamente. A imagem abaixo exemplifica como esses nós são dispostos na região de interesse.

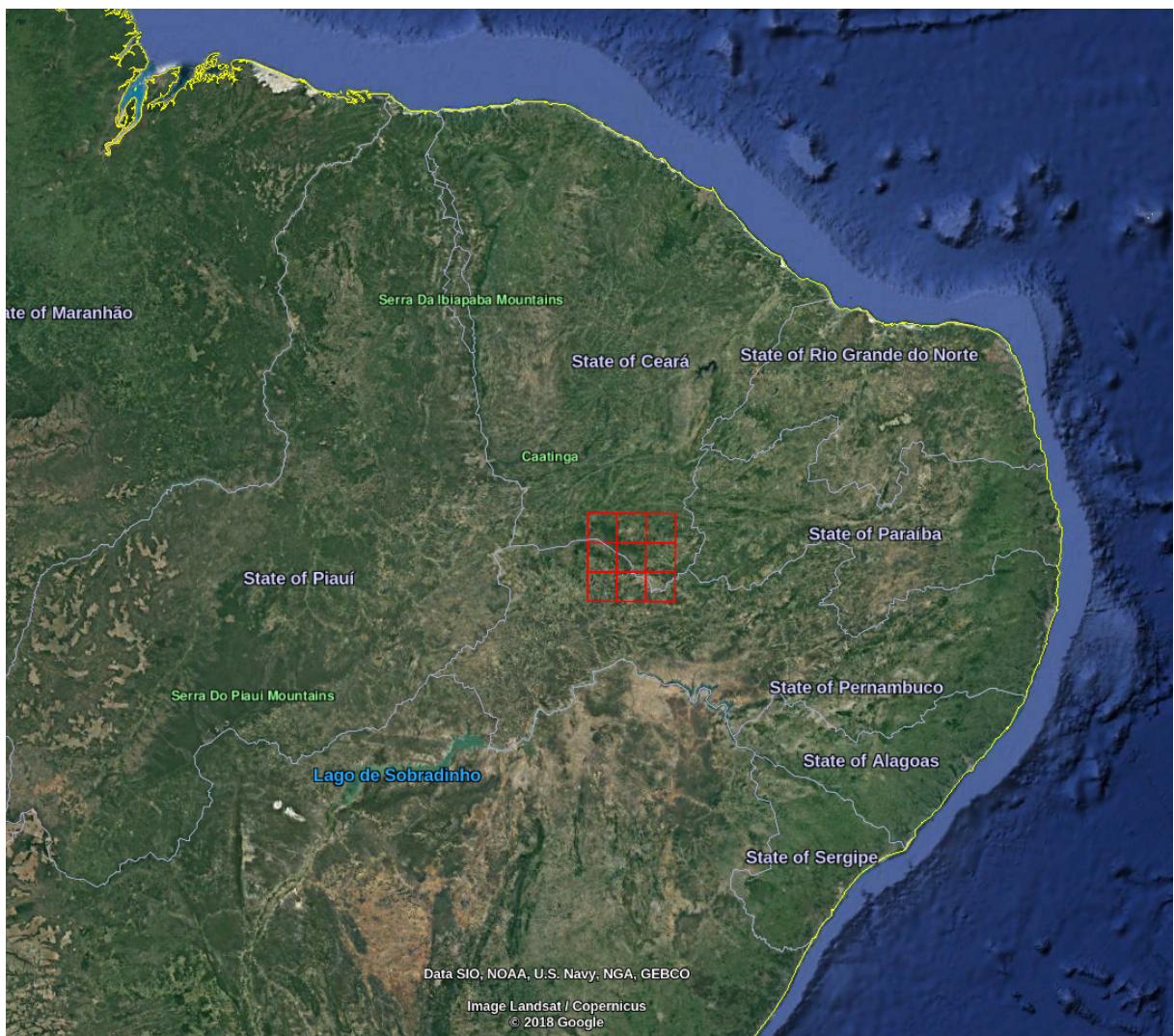


Figura 8 – Alguns nós da série ERA 5 ao sul do Ceará, Brasil.

Google earth V 7.3.2.5776. (14 de Dezembro, 2015). Rio Grande do Sul, Brasil.
 $7^{\circ} 14' 11,54''$ S, $40^{\circ} 04' 12,31''$ W, Eye alt 1808,45 km.

Cada nó cobre uma vasta região. A velocidade reportada para cada nó é uma média das velocidades da respectiva região. A campanha de medição necessária que antecede a construção de um parque eólico exige a instalação de torres de medição no local. Desse modo, a resolução dessas medições é muito maior do que a oferecida pela série ERA 5. No entanto, os dados de medições de torres não são disponibilizados publicamente. Com base na excelente correlação entre as séries ERA 5 e séries medidas por torres de projetos, acredita-se que o procedimento exposto neste trabalho possa ser aplicado com sucesso em dados medidos por torres em solo ou na nacelle dos próprios aerogeradores.

Os dados são fornecidos em base horária e compreendem o período de Janeiro de 2000 a Janeiro de 2019.

3 Caracterização dos dados

O começo de qualquer análise de séries temporais se dá pelo gráfico dos valores que assume ao longo do tempo. Por meio desse gráfico é possível identificar qualitativamente tendências, sazonalidade, ciclicidade, valores atípicos, caráter estacionário e comportamento da variância:

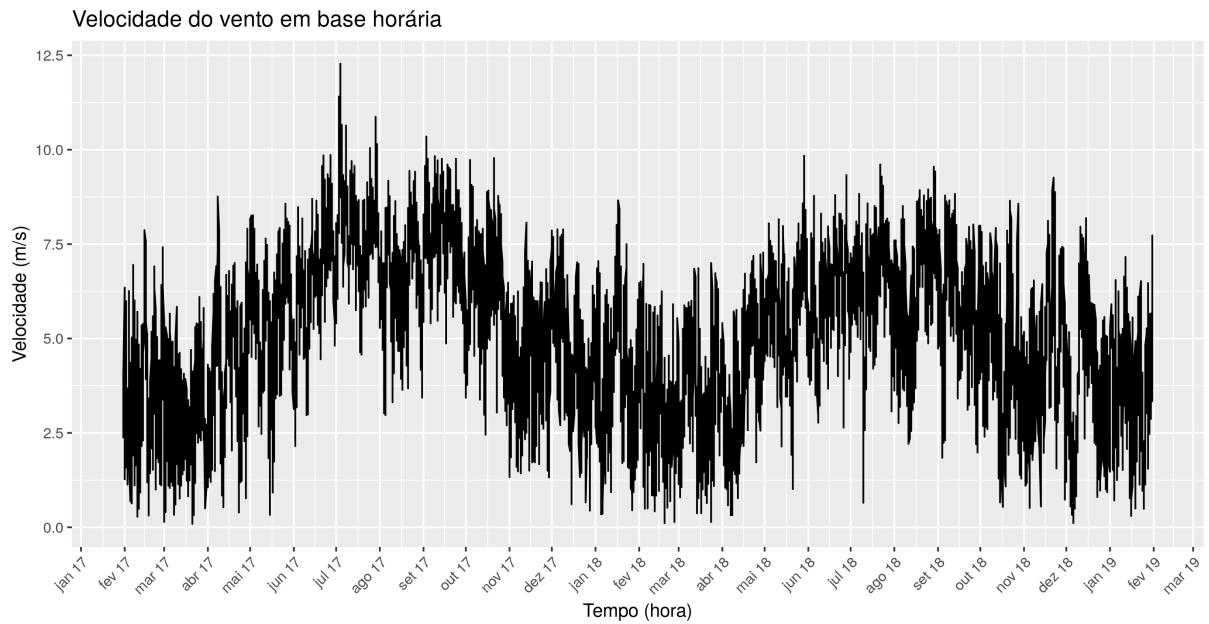


Figura 9 – Velocidade do vento registrada por satélite na região de interesse nos anos de 2017 e 2018

Apenas pela análise visual é possível perceber que a velocidade tem máximos os meses de inverno e mínimos nos de verão, que há sazonalidade anual. A série não aparenta ter derivada, oscila ao longo de um valor médio. Por apresentar sazonalidade a série não é estacionária. Como os métodos apresentados neste trabalho tomam como premissa que a série de entrada é estacionária, é necessário transformá-la previamente em uma série estacionária e ser capaz de fazer o caminho reverso após a sua previsão. Como discutido em REF isso pode ser feito por uma transformação de Box-Cox. Nessa escala não é possível afirmar algo conclusivo sobre a variância da série.

3.1 Distribuição de probabilidades

O histograma de velocidades exibido abaixo indica que a série não segue uma distribuição normal e que uma distribuição de cauda longa à direita seria mais adequada.

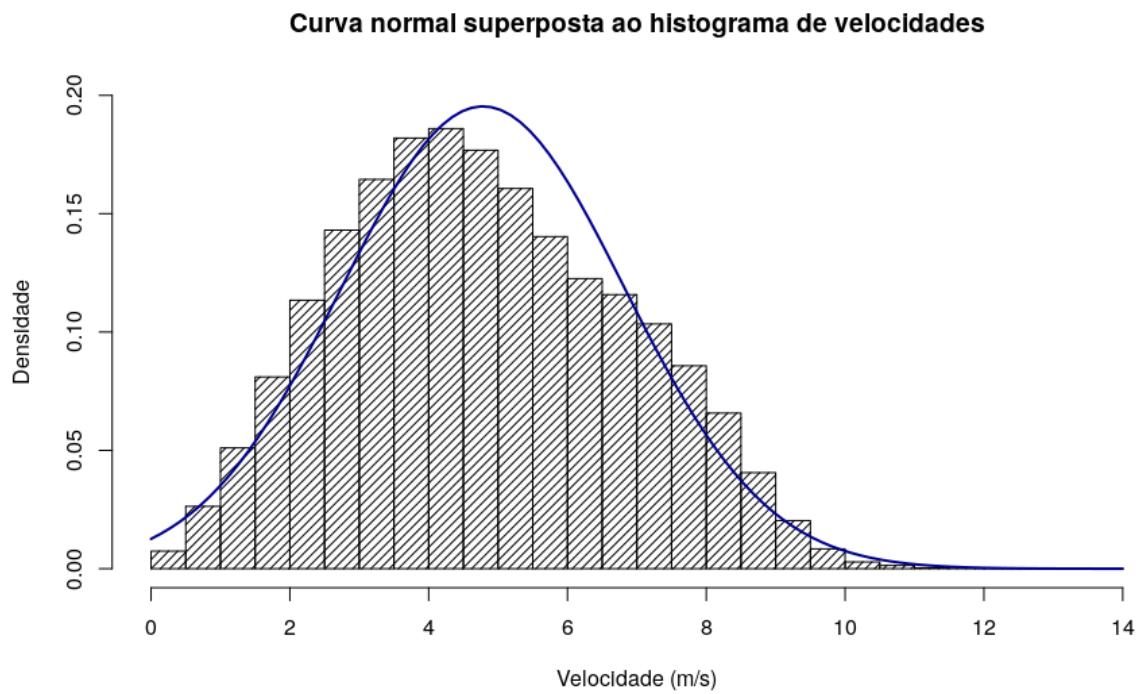


Figura 10 – Histograma de velocidades do nó noroeste da série de dados modelo.

Por meio do gráfico de Cullen e Frey é possível afirmar que a distribuição de probabilidades que melhor descreve os dados se aproxima de uma distribuição beta:

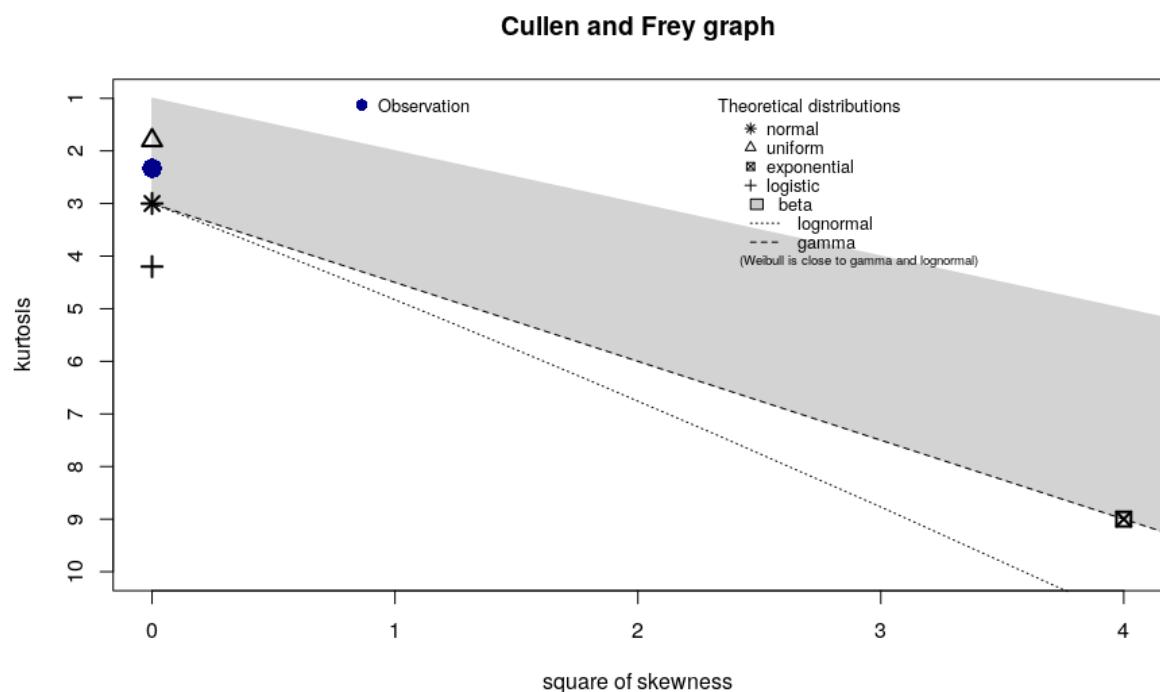


Figura 11 – Gráfico de Cullen e Frey para os dados do nó noroeste da série de dados modelo.

3.2 Sazonalidade

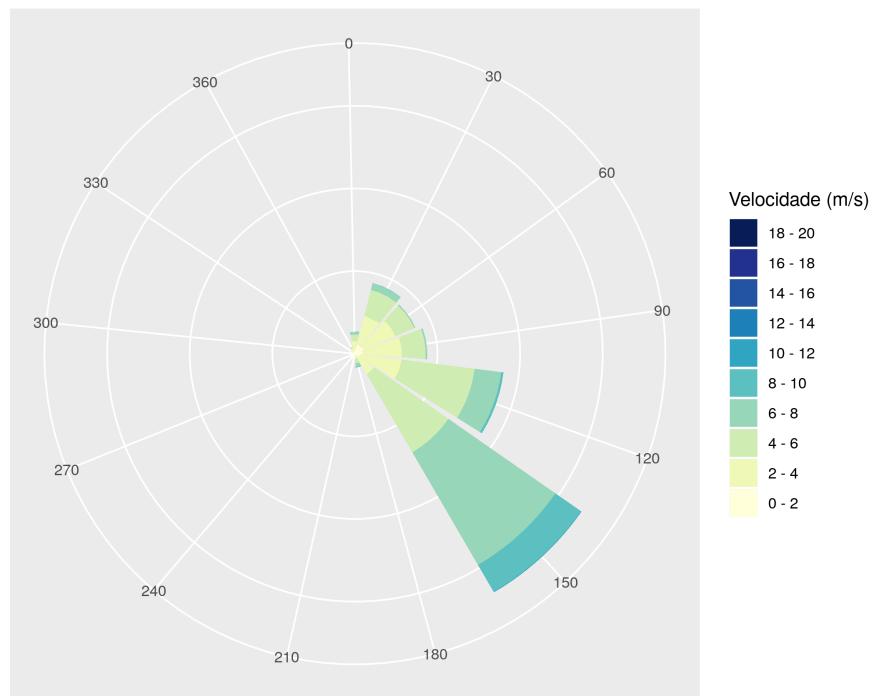


Figura 12 – Rosa dos ventos

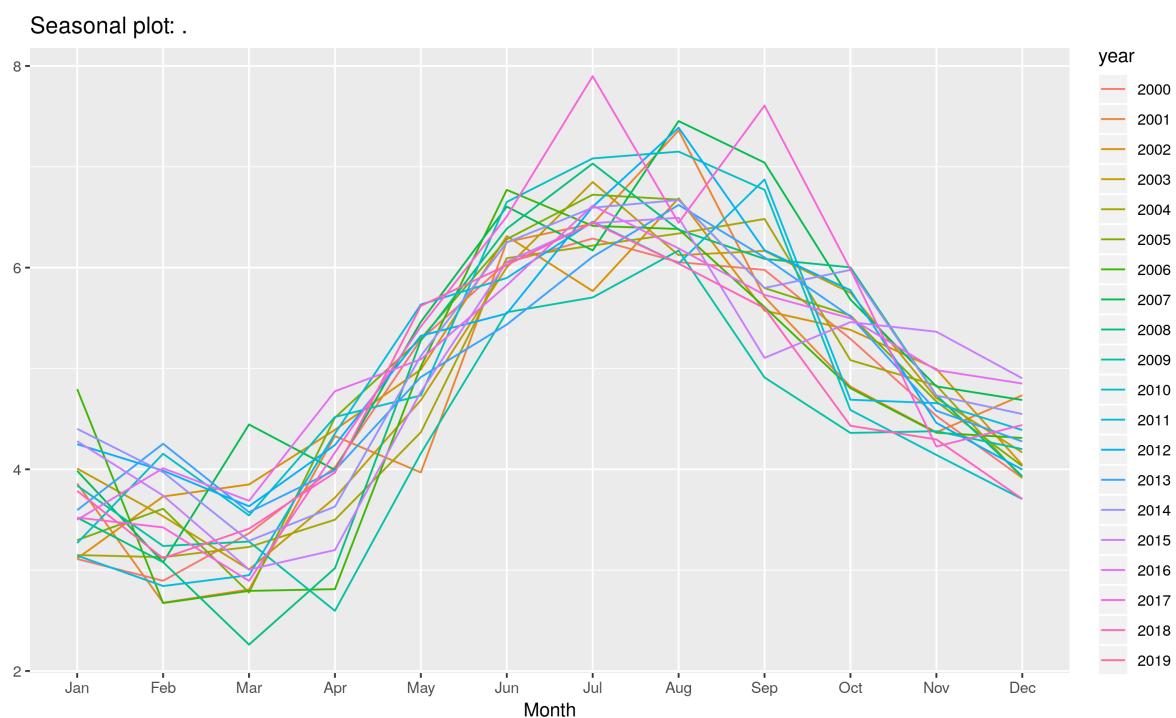


Figura 13 – Chapada

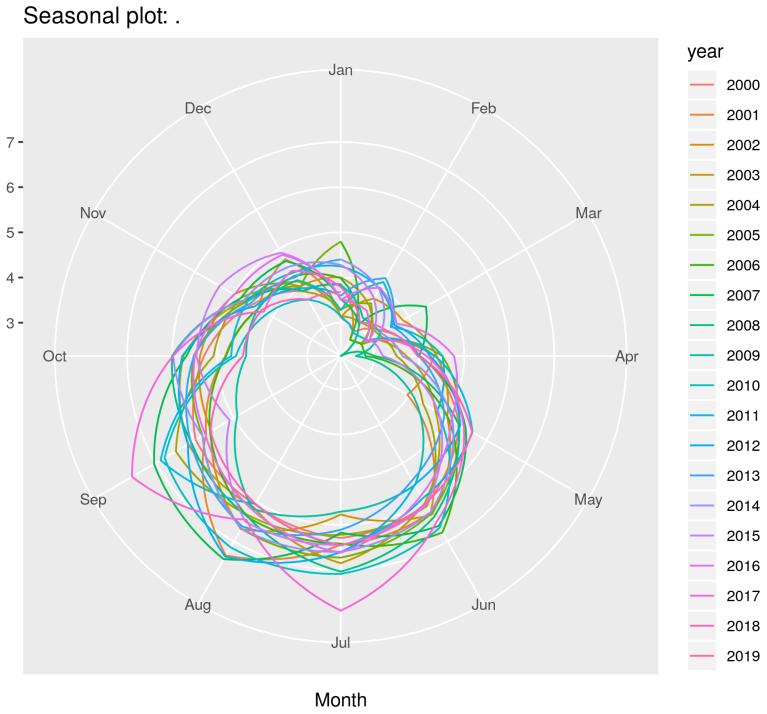


Figura 14 – Chapada

3.3 Horizonte de previsão

O horizonte de previsão é variável. Ele depende do propósito para o qual a previsão será utilizada. Alguns parques eólicos no país fazem acordos mensais sobre a energia que será entregue a rede. Nesse caso é de grande utilidade uma previsão de quanta energia será produzida no mês seguinte. A previsão minutal ou horária é, nesse caso, pouco relevante. Ao operador de subestação de distribuição de energia interessa saber quando ocorreram máximos e mínimos de produção de energia em base horária de modo que o sistema possa compensar as faltas e excessos sem perdas. Para ele a previsão minutal ou horária é essencial. Neste trabalho abordou-se tanto a previsão em escala horária quanto mensal.

Quanto mais próximo do último dado medido for a previsão mais certeza se tem do seu valor. A previsão da velocidade do vento 12 meses no futuro é muito menos confiável do que a previsão para o mês seguinte. Por esse motivo, previsões são acompanhados de um intervalo de confiança. Um intervalo de confiança de 95%, por exemplo, indica, com 95% de confiança, o intervalo de valores que velocidade medida poderá assumir.

3.4 Métodos

Existem diversos métodos para previsão de séries temporais. Ao longo do desenvolvimento deste trabalho os seguintes métodos foram considerados:

- Suavização exponencial: os valores passados contribuem para o valor atual com um peso que decai exponencialmente quanto mais distante estão do momento da previsão.
- Box-Jenkins SARIMA: um modelo robusto que faz uso de valores passados (autoregressão, AR) e erros passados (média móvel, MA), inclui integração (I) para tornar a série estacionária e estabilizar variância e conta com um processo iterativo para estimar parâmetros (Box-Jenkins).
- Rede neural de memória curta de longo prazo (LSTM): esse modelo consegue determinar com precisão o peso que deve ser dado a valores muito distantes do momento da previsão.
- Máquinas com vetor de suporte: os dados passam por uma transformação não-linear que permite a sua categorização em um espaço de dimensão elevada e posterior transformação reversa para reportar a previsão.

O método escolhido foi o Box-Jenkins SARIMA. Cada modelo assume uma série de hipóteses sobre o comportamento do que se deseja prever e possui um conjunto de parâmetros que precisam ser estimados para tal propósito. Para o modelo escolhido uma das hipóteses é de que os dados possuem autocorrelação, isto é, o valor atual depende dos n valores que o antecedem: se a velocidade foi alta um instante atrás é provável que ela seja alta agora, se no momento anterior o regime foi turbulento, é provável que no momento seguinte ele permaneça turbulento. Essa hipótese corresponde ao caráter autoregressivo (AR) do modelo. Outra hipótese do modelo é de que há autocorrelação não apenas em valores passados mas também nos erros passados. Essa hipótese corresponde ao caráter de média móvel (MA) do modelo. A interpretação dessa hipótese não é tão óbvia mas seu uso se justifica pelo fato de que um processo AR com infinitos termos pode ser descrito por um processo MA com finitos termos e vice-versa. No contexto de séries temporais o princípio da parsimônia dita que dentre os modelos que caracterizem uma série, se escolha aquele com menos parâmetros. Dessa forma a combinação de termos AR e MA permite que a estrutura da série seja capturada com poucos parâmetros.

Parte II

Implementação

Este trabalho faz uso de um modelo linear estocástico que tem como hipótese que a série temporal sob análise é gerada por uma combinação de choques aleatórios. Essa hipótese foi inicialmente proposta por Yule e Walker e extendida por diversos outros autores. Na literatura esse modelo é conhecido como modelo autoregressivo com média móvel (ARMA, do inglês autoregressive moving-average)

Um processo estocástico pode ser visto como o resultado de um filtro linear aplicado em ruído branco a_t . rodapé: Ruído branco é definido como um sinal aleatório o qual tem a mesma intensidade para qualquer frequência [fonte].

$$z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots$$

$$z_t = \mu + a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j}$$

Ruído branco tem média zero e variância constante.

Qualquer quantidade que seja observada cronologicamente é dita uma série temporal. Exemplos incluem: o preço diário das ações da Google, a quantidade mensal de chuva no Rio Grande do Sul, a produção anual de vinhos na serra gaúcha. Neste trabalho a série de interesse é de velocidade média do vento a 100 m de altura do solo no sul do estado do Ceará. Os dados são coletados por satélite e são disponibilizados em base horária no fuso horário GMT.

4 Box-Jenkins

Em 1970 George Box e Gwilym Jenkins introduziram um método empírico para abordar a previsão de séries temporais. O método ficou conhecido como método de Box-Jenkins. Ele consiste de uma série de etapas que podem ser resumidas no fluxograma abaixo:

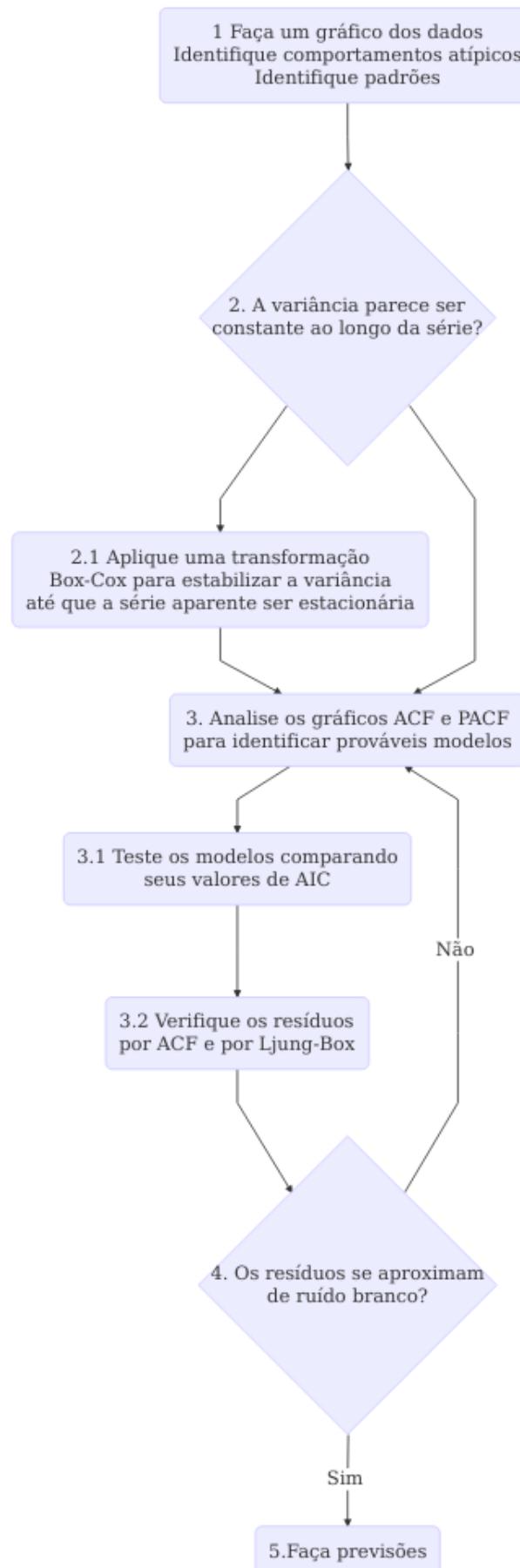


Figura 15 – O método de Box-Jenkins.

O fluxograma acima inclui algumas mudanças incorporadas por diversos outros autores tais como Hyndman & Khandakar.

Neste trabalho optou-se por essa abordagem para estimar séries temporais de velocidade do vento e seu uso será demonstrado a seguir.

4.1 Análise visual

A escolha do modelo para prever uma série temporal se dá pelas características dessa série tais como padrões, valores atípicos, tendências, sazonalidade e ciclicidade. Dessa forma o primeiro passo é visualizar a série tanto em sua forma bruta quanto sob outras formas que ressaltem outras características que não sejam visíveis na sua forma bruta. A transformação para o espaço de frequência, por exemplo, permite identificar se há periodicidade na série. Para facilitar a visualização a imagem abaixo apresenta apenas os 2 últimos anos de dados em base horária:

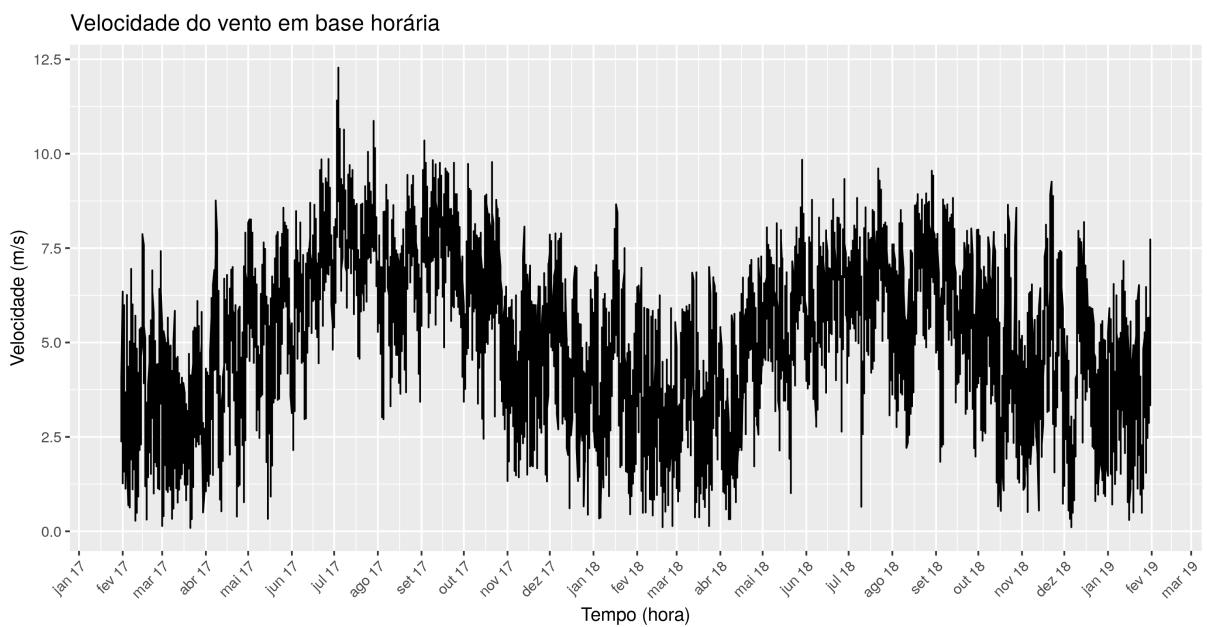


Figura 16 – Últimos dois anos da série de vento

Por meio da análise visual é possível perceber que a série não é estacionária pois em diferentes períodos a série encontra-se em diferentes níveis (mostra uma tendência, não sendo, portanto, estacionária de primeira ordem) e também apresenta sazonalidade anual (estacionária de segunda ordem).

4.2 Regime estacionário

Um processo estocástico é dito estritamente estacionário se o deslocamento da origem temporal não altera suas propriedades, ou seja, a distribuição conjunta de probabilidades calculada para o sequência de m medições z_1, z_2, \dots, z_m tomadas nos tempos t_1, t_2, \dots, t_m é a mesma àquela calculada em $t_{1+k}, t_{2+k}, \dots, t_{m+k}$ para as m medições $z_{1+k}, z_{2+k}, \dots, z_{m+k}$. $k \in \mathbb{Z}$ pode assumir tanto valores positivos quanto negativos, isto é, o deslocamento temporal pode ser tanto positivo quanto negativo. Dessa forma, um processo estacionário é caracterizado por uma distribuição de probabilidades que não varia no tempo. O modelo autoregressivo com média móvel utilizado neste trabalho toma essa propriedade como hipótese para o seu desenvolvimento, portanto, se faz fundamental garantir que a série de entrada satisfaça tal hipótese.

O primeiro passo de modo a tornar a série estacionária é remover a sazonalidade. Isso pode ser feito por uma simples diferença:

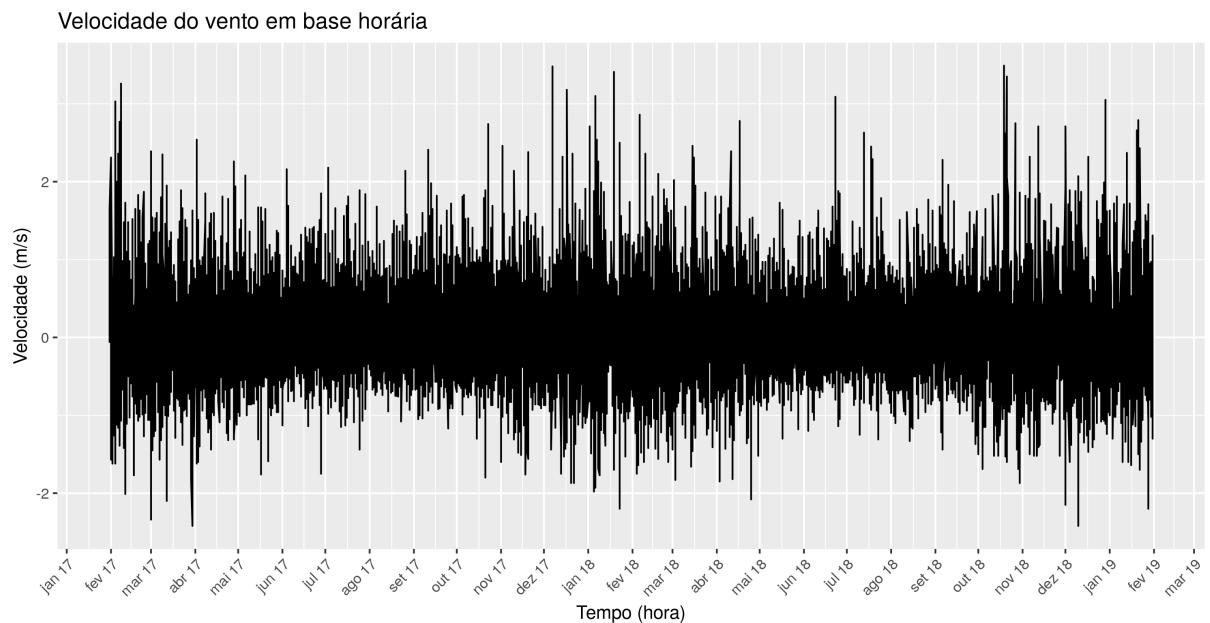


Figura 17 – Sazonalidade removida

Antes de remover a sazonalidade não é possível fazer qualquer afirmação sobre o comportamento da variância. Após removê-la por meio de uma simples diferenciação percebe-se que a variância parece ter estabilizado. No entanto essa é uma afirmação qualitativa. Pode-se obter uma confirmação quantitativa por meio de testes estatísticos chamados coletivamente como testes de raiz unitária. Optou-se por utilizar o teste Kwiatkowski-Phillips-Schmidt-Shin (KPSS) cuja hipótese nula é de que a série é estacionária (tanto em relação à tendência quanto à sazonalidade).

Aplicando o teste encontrou-se um p-valor maior que 0.1 com um nível de signifi-

cância de 0.05 de tal forma que a hipótese nula não pôde ser rejeitada. Aplicando o mesmo teste antes da remoção da sazonalidade o p-valor calculado é menor que 0.01 indicando que a série não é estacionária.

Alternativamente à simples diferença entre o valor seguinte e o valor atual, pode-se optar por uma transformação de Box-Cox. Essa transformação é definida como:

$$y_i^{(\lambda)} = \begin{cases} \sqrt[n]{\prod_{i=1}^n x_i} \ln y_i, & \text{se } \lambda = 0 \\ \frac{y_i^\lambda - 1}{\lambda \left(\sqrt[n]{\prod_{i=1}^n x_i} \right)^{\lambda-1}}, & \text{se } \lambda > 0 \end{cases} \quad (4.1)$$

O valor de lambda calculado para a série acima foi de 1.175388. A aplicação do teste KPSS indica que a série está em regime estacionário. O resultado da transformação é exibido abaixo:

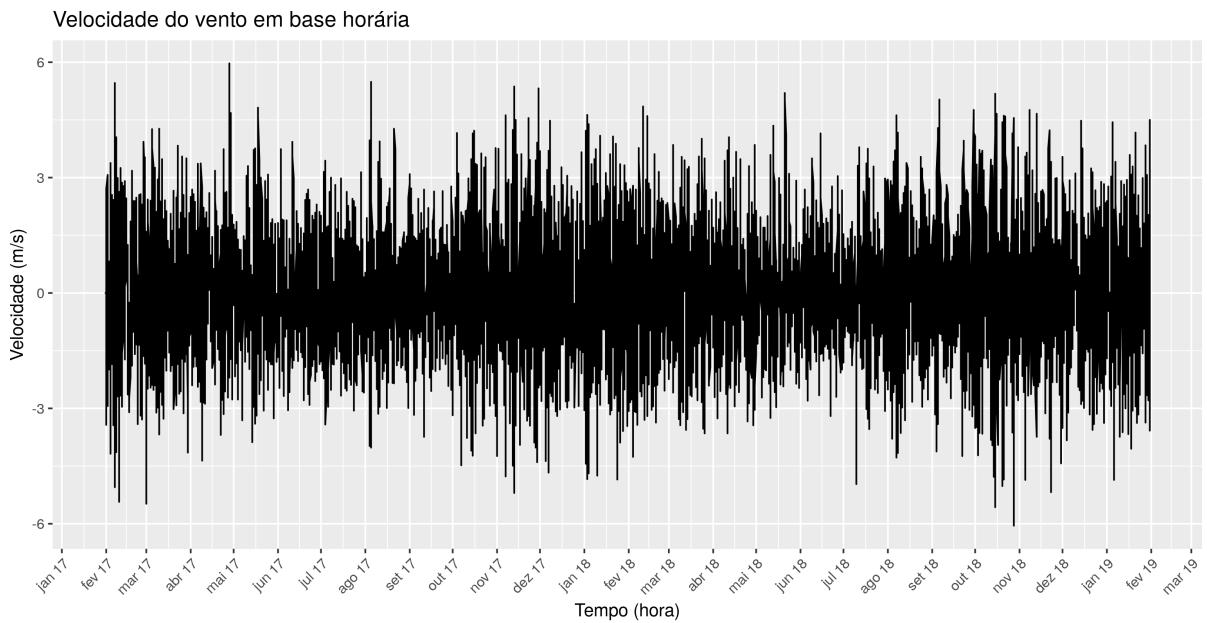


Figura 18 – Sazonalidade removida por meio de uma transformação de Box-Cox

A estabilização da variância aparente ter sido mais efetiva utilizando esse método mais sofisticado. Como resultado deste trabalho deseja-se obter um procedimento que possa realizar previsões para um período seguinte em tempo real. Essa exigência impede o uso de uma transformação Box-Cox pois devido a complexidade da operação a sua aplicação em um grande volume de dados é lenta. Além da simplicidade da diferença simples, ela produz resultados similares e ambas abordagens satisfazem testes de estacionariedade.

Dessa forma conclui-se que o parâmetro de integração (I) do modelo ARIMA que buscamos vale $d = 1$.

4.3 Autocorrelação

A etapa seguinte do método Box-Jenkins consiste em analisar os gráficos de autocorrelação e autocorrelação parcial de modo a estimar empiricamente quantos termos de autoregressão (AR) incluir e quantos de média móvel (MA).

O gráfico da função de autocorrelação (ACF) exibe o valor da correlação entre o elemento x_i da série com o elemento x_{i-k} para vários valores de k (abcissa). A autocovariância é definida por:

$$\rho_{t,t-k} = \text{Corr}(x_t, x_{t-k}) = \frac{\text{Cov}(x_t, x_{t-k})}{\sqrt{\text{Var}(x_t)\text{Var}(x_{t-k})}} = \frac{\text{Cov}(x_t, x_{t-k})}{\sqrt{\text{Cov}(x_t, x_t)\text{Cov}(x_{t-k}, x_{t-k})}} \quad (4.2)$$

Onde covariância (Cov) é definida por:

$$\text{Cov}(x, y) = E[(x - \mu_x)(y - \mu_y)] \quad (4.3)$$

Onde $E[x]$ é o valor esperado da série temporal x e μ_x seu valor médio.

Os gráficos das funções ACF e PACF são exibidos abaixo. Os termos que exibem autocorrelação significativa encontram-se acima da linha azul.

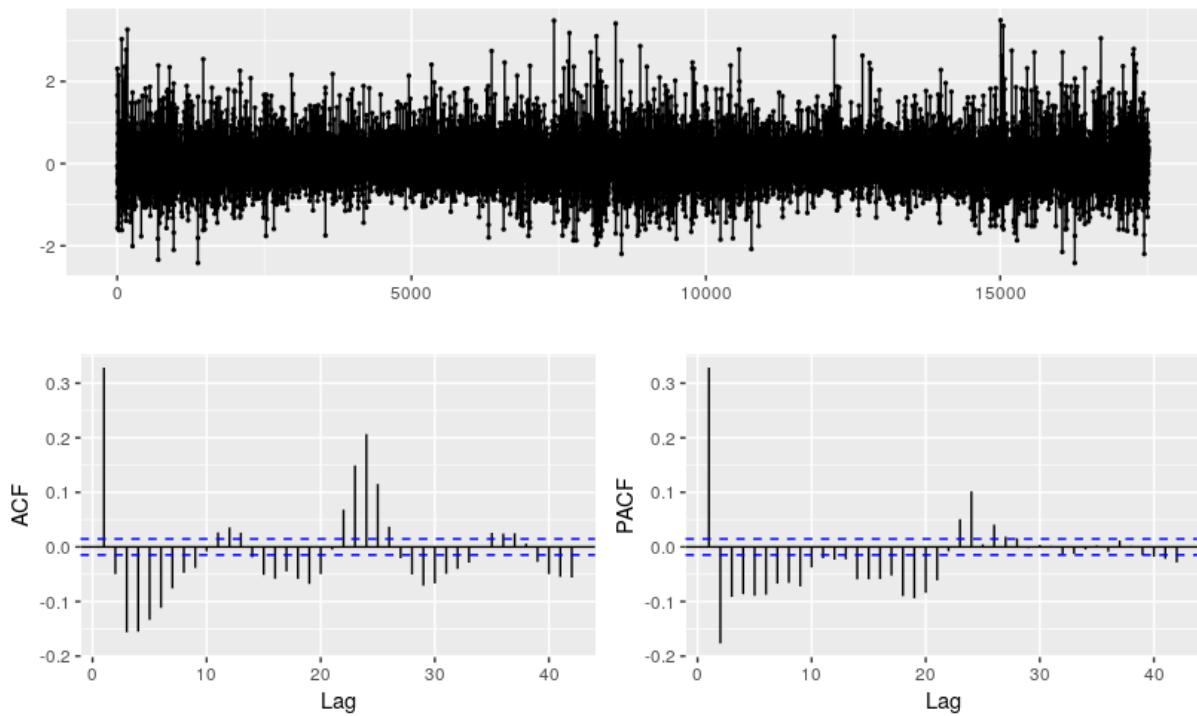


Figura 19 – Gráficos da Função de Autocorrelação (ACF) e Função de Autocorrelação Parcial (PACF)

Um pico na função ACF em $k=12$ pode significar que a série apresenta autocorrelação significativa entre um dado valor e o décimo segundo valor anterior a ele (a cada 12 horas). No entanto, isso pode ser apenas um efeito colateral de uma forte autocorrelação entre um valor e o sexto valor anterior a ele (a cada 6 horas). A função autocorrelação parcial desconta a contribuição de todos os passos calculados anteriormente. Isso é feito mensurando a autocorrelação nos resíduos do passo k após calcular a autocorrelação para o passo $k - 1$. Embora o gráfico PACF seja de mais fácil interpretação, o gráfico ACF ainda é útil porque a autocorrelação não é necessariamente linear e não pode, portanto, ser simplesmente subtraída.

4.4 Redes neurais recorrentes

Os gráficos ACF e PACF da sessão anterior indicam que a série possui memória de longo-prazo. A série exibe autocorrelação acima do nível de significância para valores da série muito distantes entre si. Observando poucos termos anteriores pode-se crer que a autocorrelação é atenuada ao longo do tempo, no entanto, quando olha-se para valores mais distantes, percebe-se que a atenuação é muita lenta para permitir um modelo ARIMA.

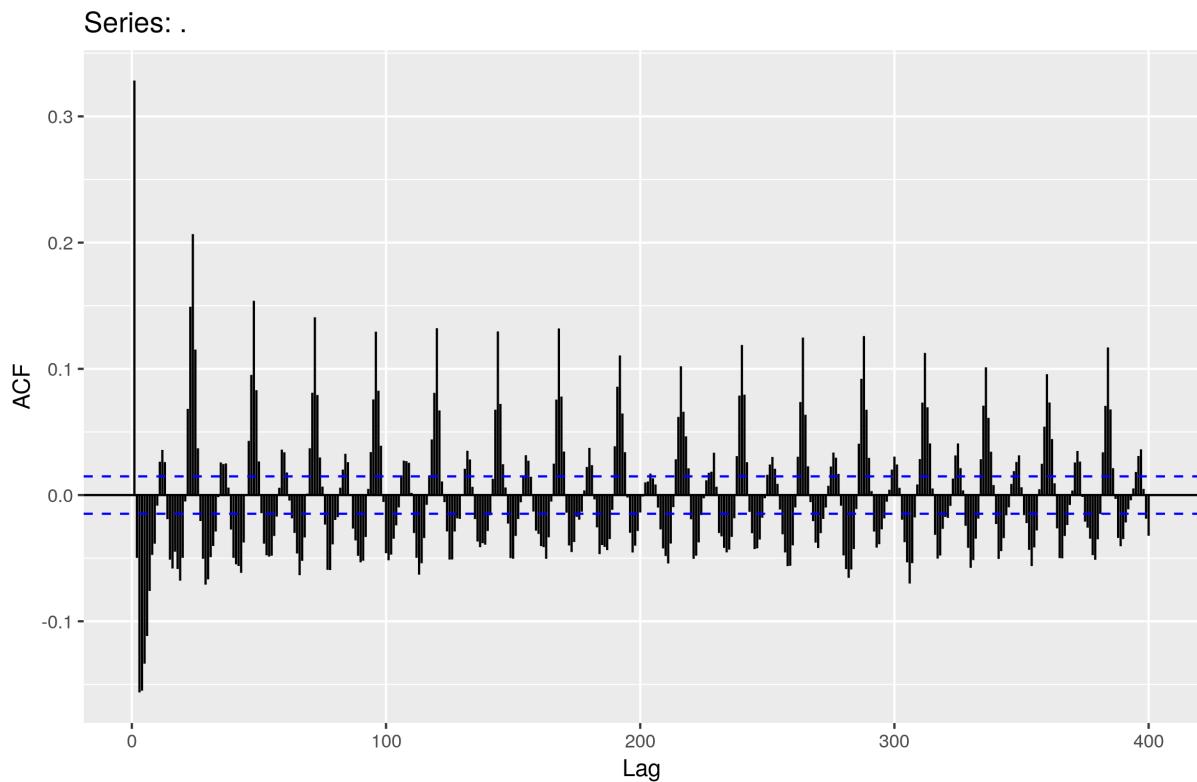


Figura 20 – Gráficos da Função de Autocorrelação (ACF) e Função de Autocorrelação Parcial (PACF)

Não é possível prosseguir naturalmente com o método de Box-Jenkins cujo próximo passo seria o teste de valores de p , q estimados a partir dos gráficos acima. Por exemplo, caso apenas a primeira e segunda componente do gráfico PACF fosse significativa, testar-se-ia um modelo ARIMA(2,1,0).

É necessário uma mudança de abordagem. Duas tentativas foram feitas. A primeira consistiu em abandonar modelos ARIMA e optar por uma modelo de rede neural recorrente. O modelo mais adequado é o modelo LSTM (Long Short-Term Memory) o qual consegue levar em conta a existência de memória de longo prazo na série e utilizar na previsão de curto prazo. Experimentou-se com vários modelos e escolhido contou com 1 camada de entrada, 1 camada oculta com 4 blocos LSTM com uma camada de saída, todas intercaladas com uma camada de descarte para evitar aprendizado em excesso (overfitting). Apesar do esforço empregado, não foi possível encontrar parâmetros adequados para o modelo. O resultado da previsão, mostrado abaixo, foi insatisfatório. Acredita-se que um maior esforço no entendimento e devida configuração pudesse trazer melhores resultados, no entanto, optou-se por abandonar esse modelo devido ao longo tempo de treinamento, isto é, o tempo em que o código leva para olhar para os valores passados da série e estimar os parâmetros utilizado na previsão de valores futuros.



Figura 21 – Modelo LSTM aplicado a série. Velocidade no eixo y. Tempo em horas no eixo x.

4.5 Janela de dados

A segunda abordagem para contornar o problema de memória longa da série consiste em restringir a janela de dados. Por um lado a restrição do intervalo de análise da série resolve o problema de memória de longo-prazo (não há mais longo-prazo). Por outro lado, os parâmetros estimados para um janela de dados pode não ser válido para a janela seguinte, exigindo, dessa forma, o cálculo constante dos novos parâmetros a medida que a janela cobre toda a série. Se o tempo para o cálculo desses parâmetros for menor que o tempo de aquisição dos dados esse torna-se apenas uma questão de ter hardware dedicado a esse cálculo ininterruptamente. Essa abordagem também permite que mudanças na série sejam facilmente incorporadas ao modelo, tal como, por exemplo, trocar a série horária por uma série de 10 minutos ou uma mudança natural de tendência na série devido, por exemplo, ao aumento ou diminuição da temperatura da região.

Dessa forma, decidiu-se por restringir a janela de dados a semanas ao invés de anos. A última semana de dados é exibida abaixo. Aplicando o teste KPSS percebe-se que é necessária uma diferenciação an para torná-la estacionária.

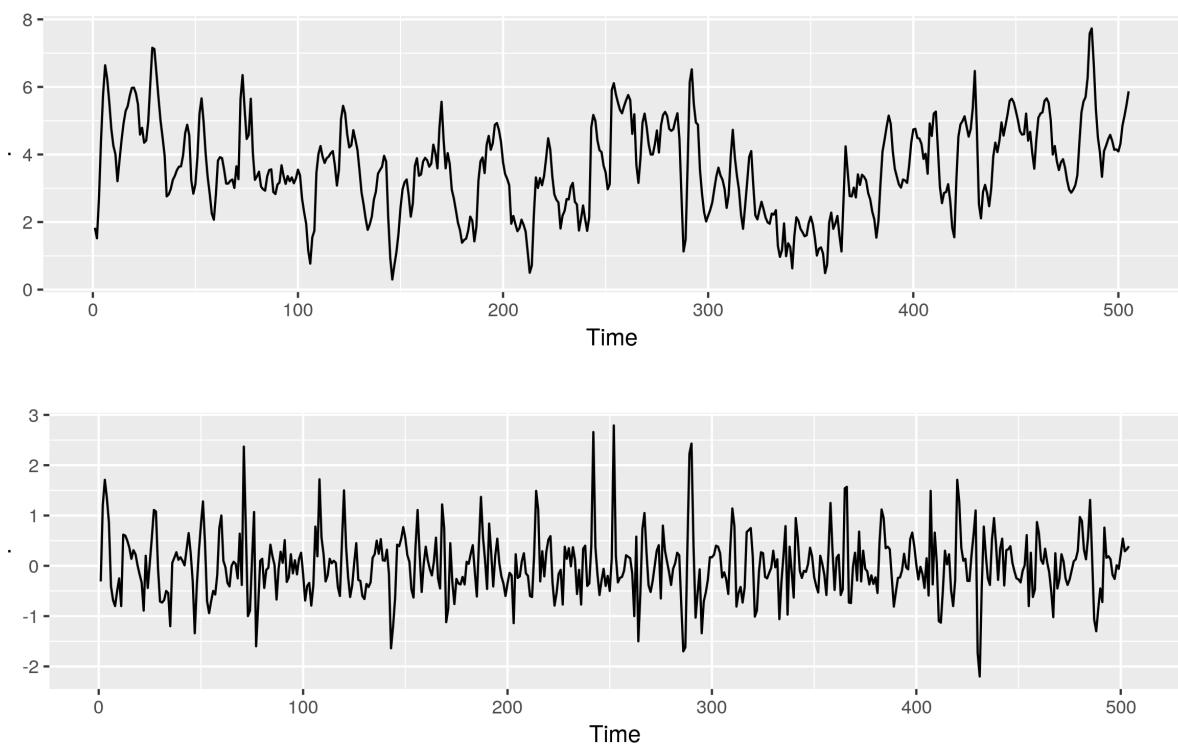


Figura 22 – 3 últimas semanas de dados da série. Antes da diferenciação (superior) e após (inferior).

Analizando os gráficos das funções de autocorrelação é possível elencar alguns possíveis modelos ARIMA. O gráfico da função ACF indica quantos termos MA incluir enquanto a função PACF indica quantos termos AR incluir. De acordo com a ACF a influência de autocorrelação se estende apenas até o sexto termo . No entanto valor respectivo está próximo do nível de significância e não é tão simples satisfazer as condições de convergência quando se inclui muitos termos MA, portanto, é interessante testar um modelo MA(3). A função PACF apresenta um valor significativo para o termo imediatamente anterior sugerindo um modelo AR(1).

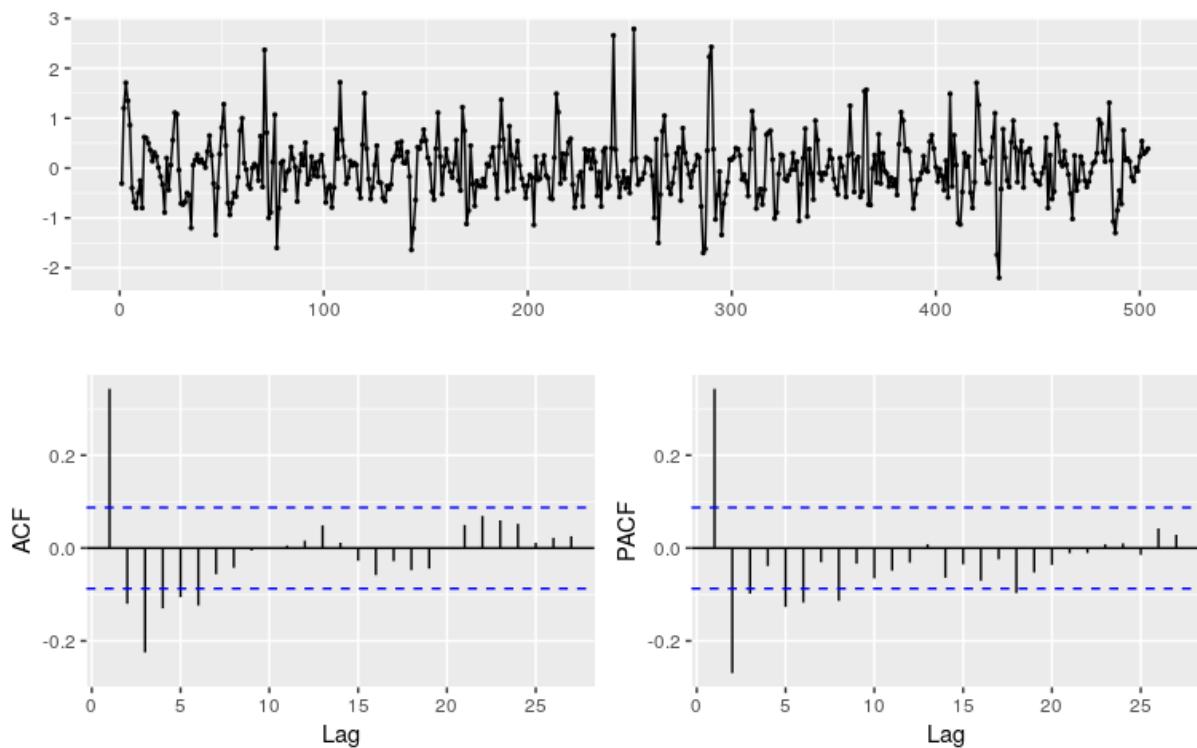


Figura 23 – Gráficos da Função de Autocorrelação e Autocorrelação Parical para 2 semanas de dados da série.

Após essas observações empíricas pode-se testar quantitativamente o desempenho de alguns possíveis modelos:

- ARIMA(0,1,1)
- ARIMA(1,1,1)
- ARIMA(1,1,2)
- ARIMA(1,1,3)

Antes de fazer previsões é necessário estabelecer critérios para comparar o desempenho de diferentes modelos.

4.6 Medidas de qualidade de previsões

Para prever e quantificar a qualidade da previsão de uma série temporal é necessário dividi-la em um conjunto de treinamento e outro de teste. A divisão pode ser, por exemplo, 80% dos dados destinados ao conjunto de treinamento e os 20% subsequentes ao conjunto de teste. Empiricamente é desejável que o conjunto de teste seja igual ao tamanho

do horizonte de previsão que se deseja obter. O conjunto de treinamento é utilizado para computar os parâmetros do modelo enquanto que o conjunto de teste é utilizado para quantificar o quanto bem o modelo é capaz de prever os dados subsequentes. Existem diversas maneiras de mensurar a qualidade das previsões geradas por um modelo. Deve-se evitar usar os resíduos como medida pois eles são resultado da estimativa dos parâmetros do modelo aos dados de treinamento. Como qualquer série pode ser aproximada por um modelo com infinitos parâmetros, os resíduos não dão informação sobre o poder preditivo do modelo para valores futuros, desconhecidos. A

O erro de previsão é dado pela diferença entre o valor medido, y_{t+h} , e o valor previsto pelo modelo, \hat{y}_{t+h} :

$$e_{t+h} = y_{t+h} - \hat{y}_{t+h}$$

Existem muitas maneiras de quantificar o erro de previsão. As seções seguintes abordam as características de cada uma e identificam as que serão usadas para quantificar os resultados deste trabalho.

4.6.1 Erro médio absoluto (MAE)

O erro absoluto médio é dado pela média do módulo do erro de previsão. Essa medida penaliza o resultado tanto por erros negativos quanto positivos. A minimização dessa medida de erro resulta em estimativas que se aproximam da mediana dos dados.

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^T |e_t|$$

4.6.2 Raíz do quadrado da média do erro (RMSE)

A raíz do quadrado da média do erro é dada por:

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T e_t^2}$$

A minimização dessa medida de erro resulta em estimativas que se aproximam da média dos dados.

4.6.3 Erro absoluto médio escalonado (MASE)

Ao contrário das medidas anteriores, essa possui a vantagem de que é independente da escala dos dados medidos:

$$\text{MAPE} = \frac{\frac{1}{T} \sum_{t=1}^T |e_t|}{\frac{1}{T-1} \sum_{t=2}^T |y_t - y_{t-1}|}$$

4.6.4 Validação cruzada

A validação cruzada consiste em usar uma janela de treinamento para calcular o erro de previsão para a próxima medida, expandir a janela em uma medida e calcular o novo erro e assim por diante sob um conjunto de teste. A medida de erro é dada pela média dos erros individuais. Qualquer uma das medidas acima pode ser utilizada no cálculo da validação cruzada. Essa é a maneira mais robusta de quantificar o erro de previsão pois atua efetivamente em dados que não foram utilizados para construir o modelo. A idéia da medida pode ser ilustrada da seguinte forma:

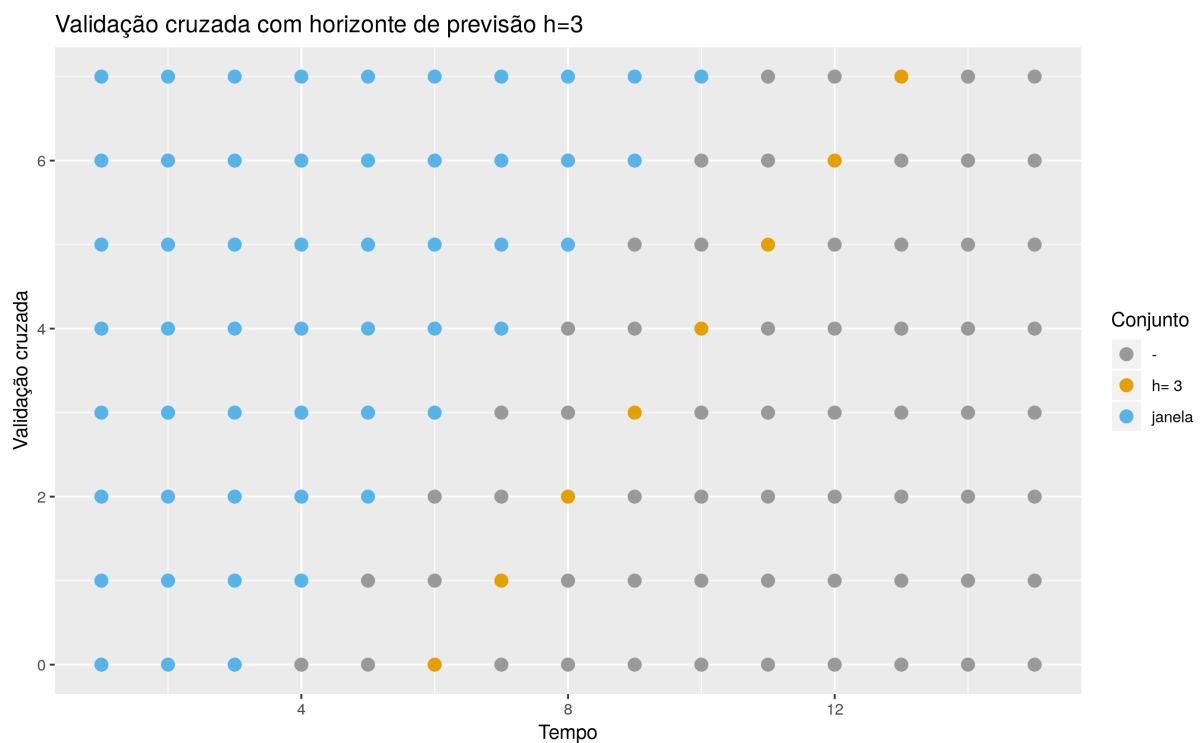


Figura 24 – Chapada

Onde as medidas de treinamento estão em azul e as de teste em amarelo. Na linha de baixo, mensura-se o erro de previsão que resulta ao tentar prever três horas a frente da hora atual. A linha acima inclui a medida seguinte ao conjunto de treinamento e tenta prever, novamente, três horas a frente. Isso é feito várias vezes. Os erros resultantes são calculados, por exemplo, por MASE e a sua média é reportada como valor final. Em todos os testes de modelos que seguem as métricas reportadas são calculadas dessa forma.

4.7 Estabilidade

Os parâmetros p, d, q de modelos ARIMA não podem assumir quaisquer valores. Existem critérios de existência e estabilidade que basicamente garantem a convergência do polinômio que representam.

Um modelo ARIMA pode ser escrito da seguinte forma:

$$\phi(B)(1 - B)^d y_t = c + \theta(B)\varepsilon_t \quad (4.4)$$

Sendo d o grau de diferenciação aplicado, $\phi(B)$ é o polinômio de ordem p :

$$\phi(B) = (1 - \phi_1 B - \cdots - \phi_p B^p) \quad (4.5)$$

$\theta(B)$ é o polinômio de ordem q :

$$\theta(B) = (1 + \theta_1 B + \cdots + \theta_q B^q) \quad (4.6)$$

B o operador definido por:

$$By_t = y_{t-1} \quad (4.7)$$

A condição de regime estacionário de qualquer modelo ARIMA é de que as raízes complexas do polinômio $\phi(B)$ estejam fora do círculo unitário. Enquanto que a condição de invertibilidade é de que as raízes complexas de $\theta(B)$ estejam fora do círculo unitário. Pode-se verificar ambas as condições por meio do gráfico das raízes desses polinômios. Para facilitar a visualização exibe-se o gráfico do respectivo polinômio inverso de tal forma que as condições sejam satisfeitas se as raízes estiverem dentro do círculo unitário.

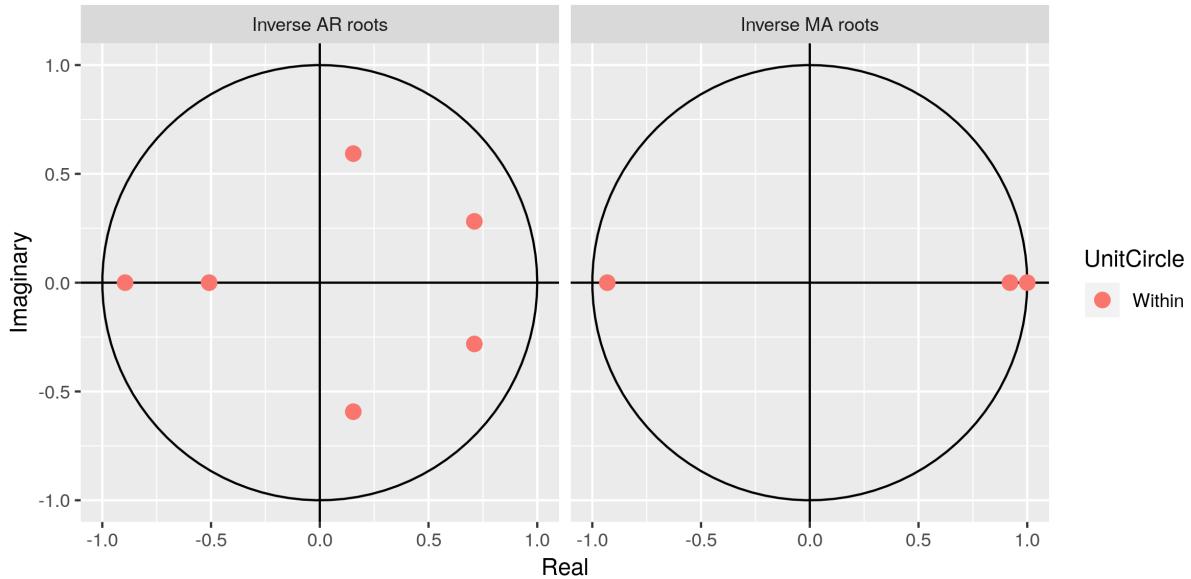


Figura 25 – Condições de regime estacionário e invertibilidade para um modelo ARIMA(6,1,0).

4.8 Previsão

A estabilidade dos modelos propostos na seção 4.5 foi verificado e todas as raízes estão fora do círculo unitário. Dessa forma pode-se comparar o desempenho de cada modelo atuando na série de dados. Para isso tomou-se 2 semanas anteriores como conjunto de treino para estimar os parâmetros de cada modelo e a semana seguinte como conjunto de teste para comparar o que os modelos predizem com o que efetivamente foi registrado.

Visualmente a previsão pode ser demonstrada superpondo a curva prevista aos dados medidos. Esses modelos permitem que se calculem também intervalos de confiança. Optou-se por exibir os intervalos de 80% e 95% em azul escuro e claro respectivamente. Um intervalo de 95% indica que o modelo prevê com 95% de confiança que o valor medido estará entre os limites indicados.

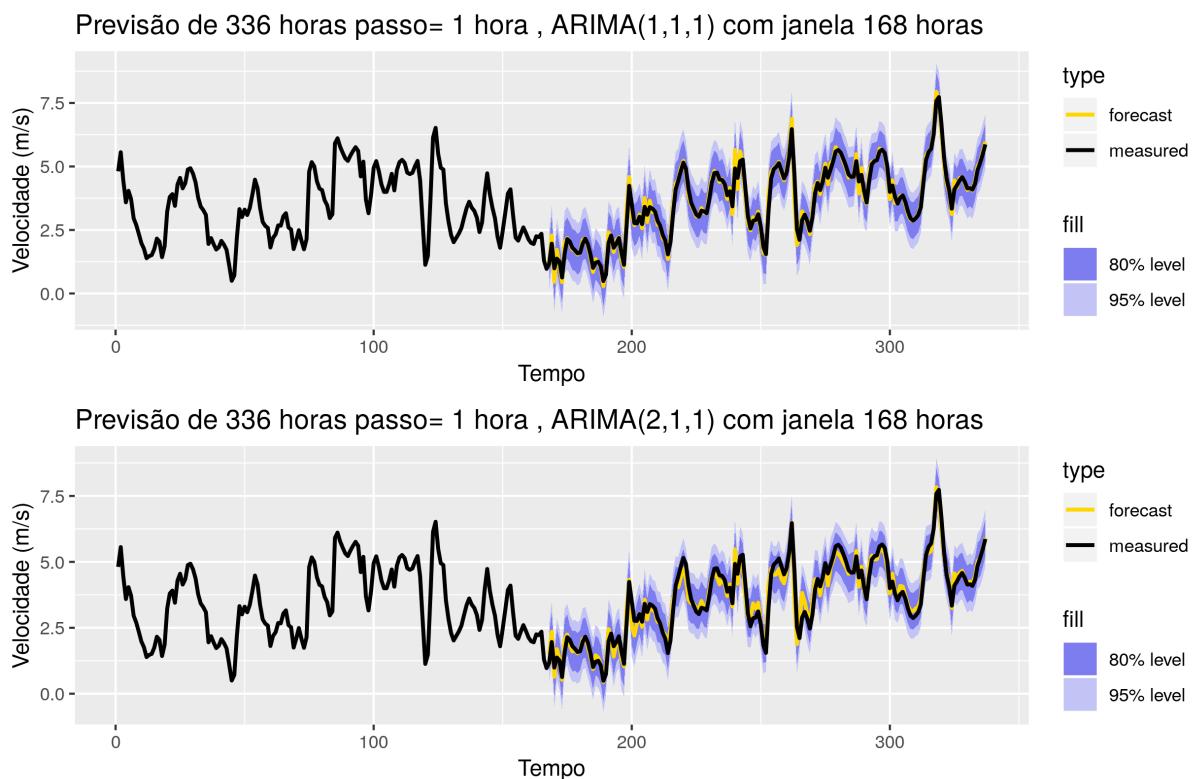


Figura 26 – Condições de regime estacionário e invertibilidade para um modelo ARIMA(6,1,0).

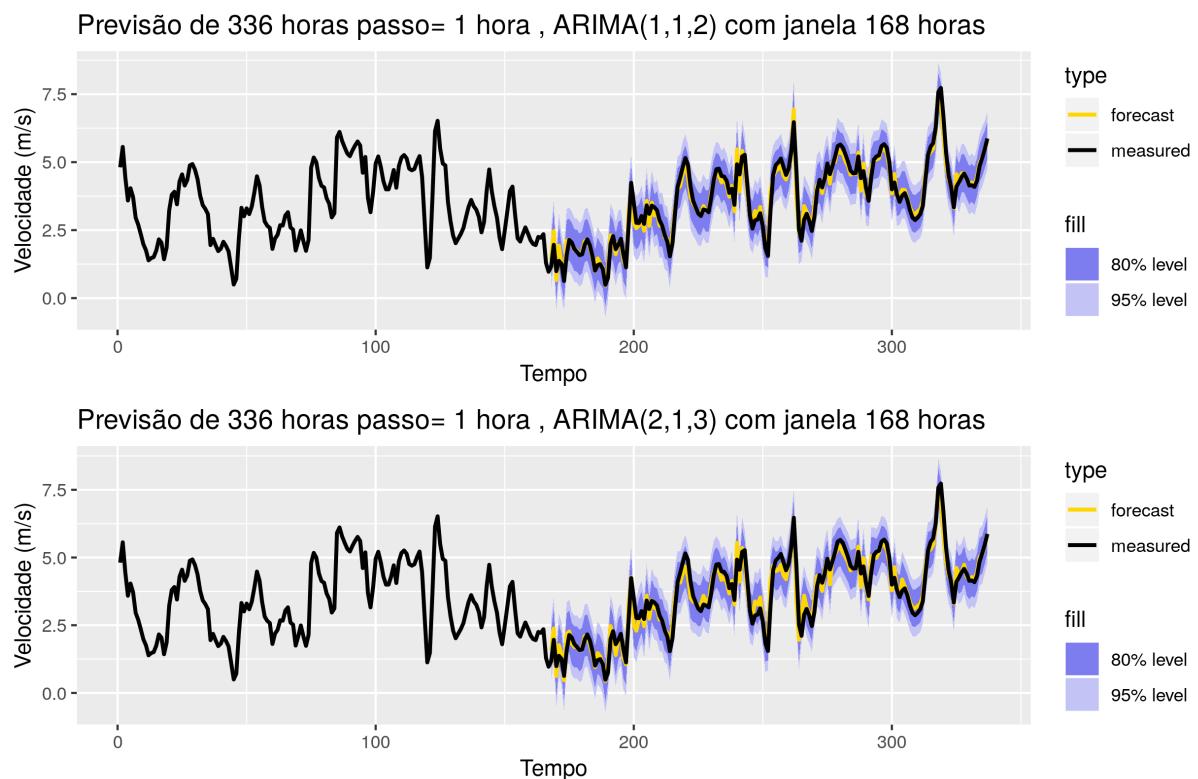


Figura 27 – Condições de regime estacionário e invertibilidade para um modelo ARIMA(6,1,0).

Os valores quantitativas das diversas medidas de erro introduzidas na seção 4.6, usando o procedimento de validação cruzada descrito em 4.6.4.

ARIMA(p,d,q)	ME	RMSE	MAE	MPE	MAPE
1,1,1	-0.007948941	0.2231684	0.161981	0.3817872	5.731274
2,1,1	0.00824027	0.2973698	0.2285103	-0.7558966	7.622429
1,1,2	0.01559968	0.2490103	0.1870849	-0.8393794	6.151953
2,1,3	0.01455176	0.3189862	0.2380204	-0.8382305	7.968127

A minimização das métricas RMSE, MAE, MPE e MAPE todas sugerem que o modelo ARIMA(1,1,1) apresenta melhor desempenho para essa janela de dados. No entanto observa-se que esse modelo não fornece sempre os melhores resultados pra toda a extensão da série tampouco isso era esperado. A seção seguinte propõe um método de estimar continuamente os melhores parâmetros ARIMA para uma dada janela de dados.

5 Modelo Autoregressivo Variável

Ao longo do estudo observou-se que para diferentes períodos da série de dados, diferentes modelos ARIMA apresentavam melhor potencial preditivo não sendo possível elencar um único modelo que capturasse a estrutura da série de forma universal. Duas soluções para esse problema foram exploradas. Uma delas é optar por outro modelo. A outra, abordada nessa seção, é permitir que os parâmetros ARIMA fossem adaptativos, sendo recalculados, a medida que mais dados são coletados. Essa é uma estratégia interessante por vários motivos. É possível estimar os parâmetros de um modelo ARIMA rapidamente de forma que a previsão esteja disponível ininterruptamente mesmo em escalas temporais da ordem de minutos. Qualquer mudança estrutural na grandeza física, tal como devido ao aumento da temperatura média do planeta, é automaticamente capturada pelo modelo.

A imagem abaixo exibe a evolução temporal dos graus de liberdade do modelo em um espaço de fase sendo possível entender como os parâmetros variam ao longo do tempo e consequentemente como a estrutura dos dados (limitada pela representação por um modelo ARIMA) varia. O tempo é representado pela escala em cor. O parâmetro de diferenciação, d , do modelo assume durante quase todo o período o valor 1 e portanto exibe-se apenas os parâmetros p e q .

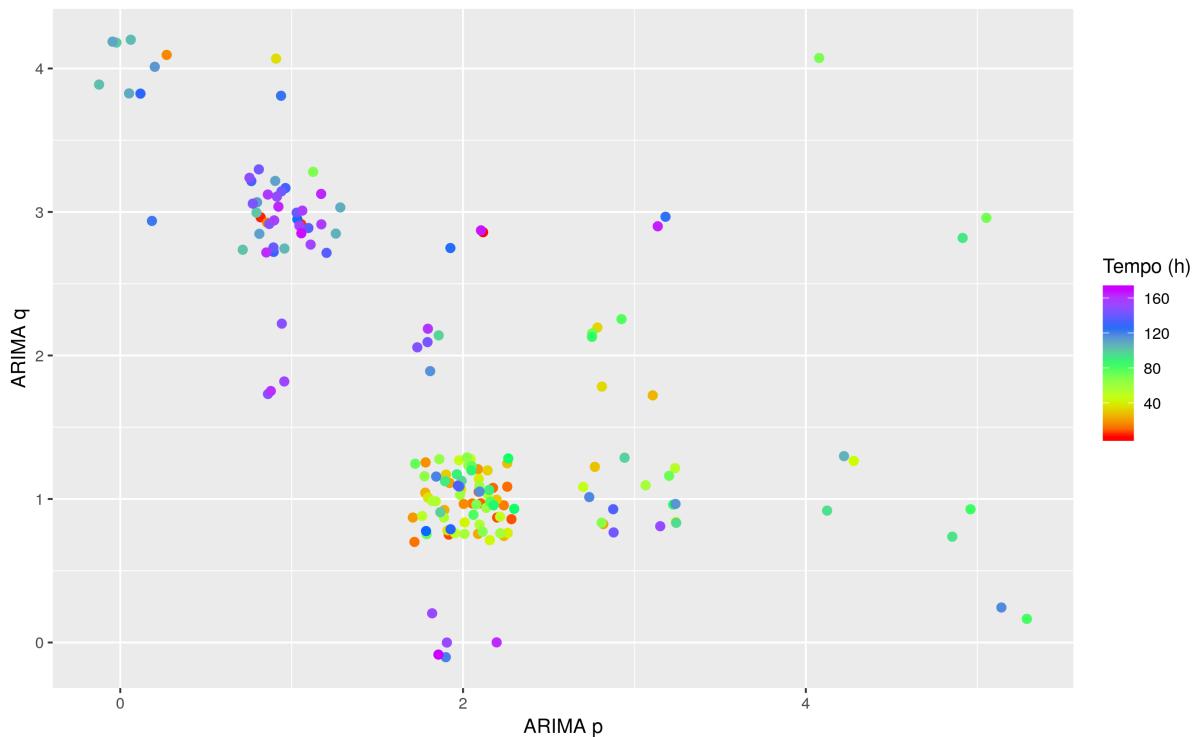


Figura 28 – Evolução dos parâmetros p e q de um modelo ARIMA ao longo do tempo.

Modelos ARIMA são definidos pelos parâmetros p,d,q . O parâmetro p representa o número de termos anteriores da série que influenciam o valor atual. Os pesos com que cada termo anterior contribui para a estimativa do valor seguinte são parâmetros a serem estimados por regressão linear. O parâmetro q tem um papel análogo ao do parâmetro q . Ele refere-se à inclusão dos q erros anteriores na estimativa do valor seguinte.

Os parâmetros assumem apenas valores inteiros. Para facilitar a visualização da evolução temporal adicionou-se um ruído aleatório a cada ponto de modo a evitar a sua sobreposição. Fica claro que no início do período de 160 horas o principal modelo ARIMA é o ARIMA(2,1,1) enquanto que ao final do período o modelo preferencial é o ARIMA(1,1,3). Também observa-se que diversos outros modelos são assumidos com menor frequência.

Apesar do apelo prático, o método não se mostrou satisfatório:

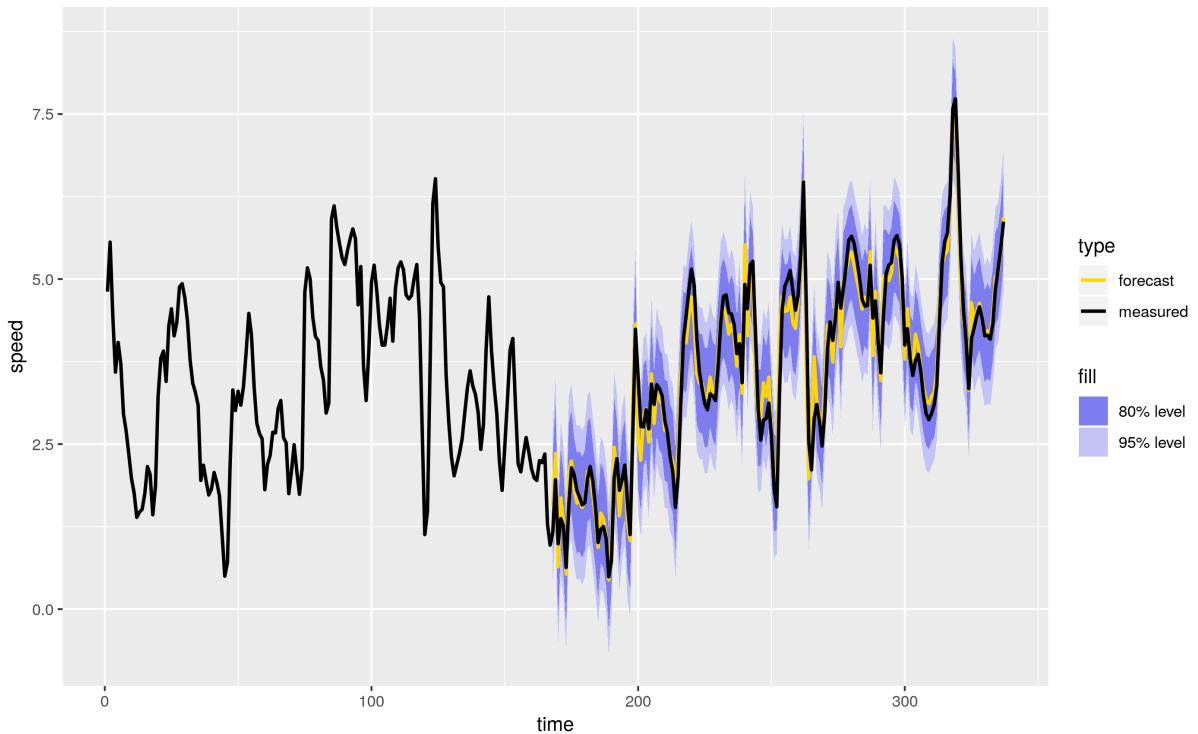


Figura 29 – Evolução dos parâmetros p e q de um modelo ARIMA ao longo do tempo.

6 GARCH

Séries de dados de vento possuem uma característica fundamental em comum que não foi modelada até então: sua variância não é constante. Os esforços dispendidos foram no sentido de neutralizar a variância. Além disso o comportamento da variância em séries de velocidade do vento é mais simples, ao contrário de séries financeiras.

6.1 Base horária

Em base horária os períodos de maior e menor variância são facilmente previsíveis e possuem uma interpretação física clara. Períodos de variância elevada são observados durante períodos de simultânea alta temperatura do ar, o que o torna mais turbulento e portanto aumenta a variabilidade das medições, enquanto que, à noite por exemplo a atmosfera próxima ao solo se estratifica em camadas bem definidas e comportadas em um regime laminar, longe da turbulência e de tal forma apresentam baixa variabilidade.

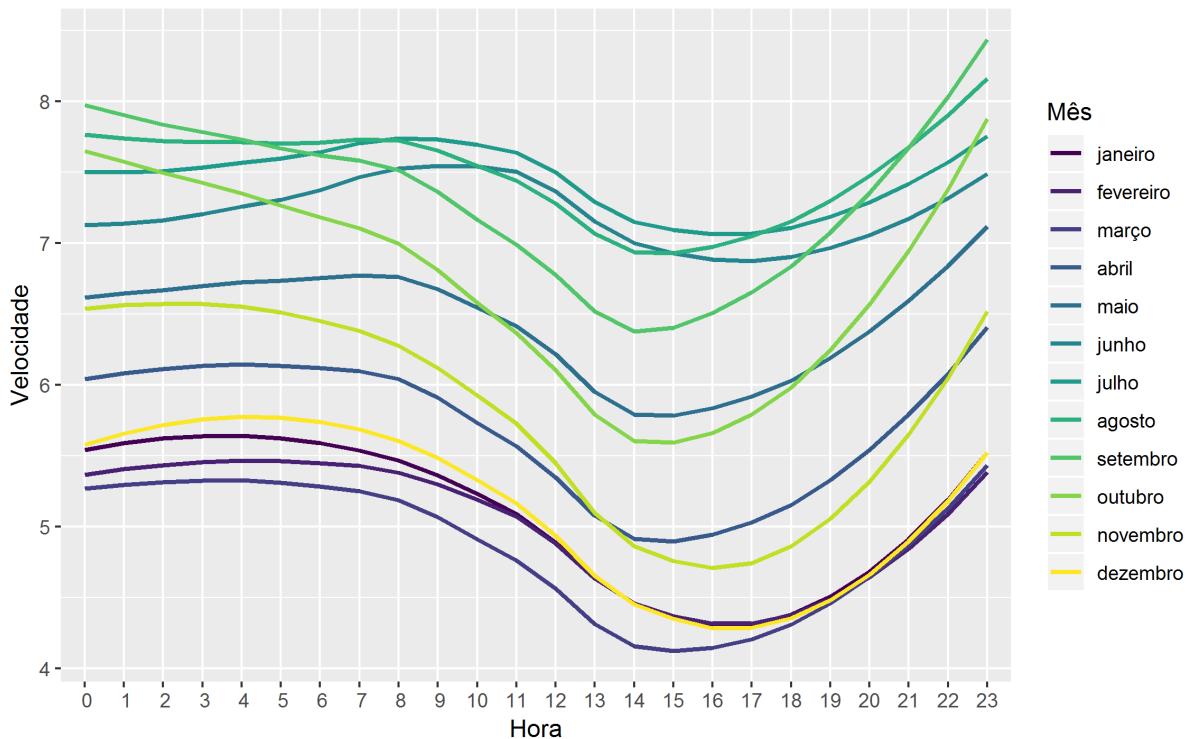


Figura 30 – Evolução dos parâmetros p e q de um modelo ARIMA ao longo do tempo.

O modelo GARCH estende o modelo ARIMA de modo a modelar a variação na variância da série temporal. Este modelo é um acrônimo que significa heterocedasticidade

(variância) condicional autoregressiva generalizada. Em um modelo GARCH a variância é modelada por:

$$\sigma_t^2 = \underbrace{\beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2}_{\text{autoregressão}} + \underbrace{\omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2}_{\text{média móvel}} \quad (6.1)$$

$$\sigma_t^2 = \sum_{i=1}^q \beta_i \sigma_{t-i}^2 + \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 \quad (6.2)$$

A equação 6.1 deixa claro a analogia entre um modelo ARIMA e um GARCH. ARIMA aplica um filtro na própria série enquanto GARCH aplica na sua variância. Os resultados desse modelo são exibidos abaixo:

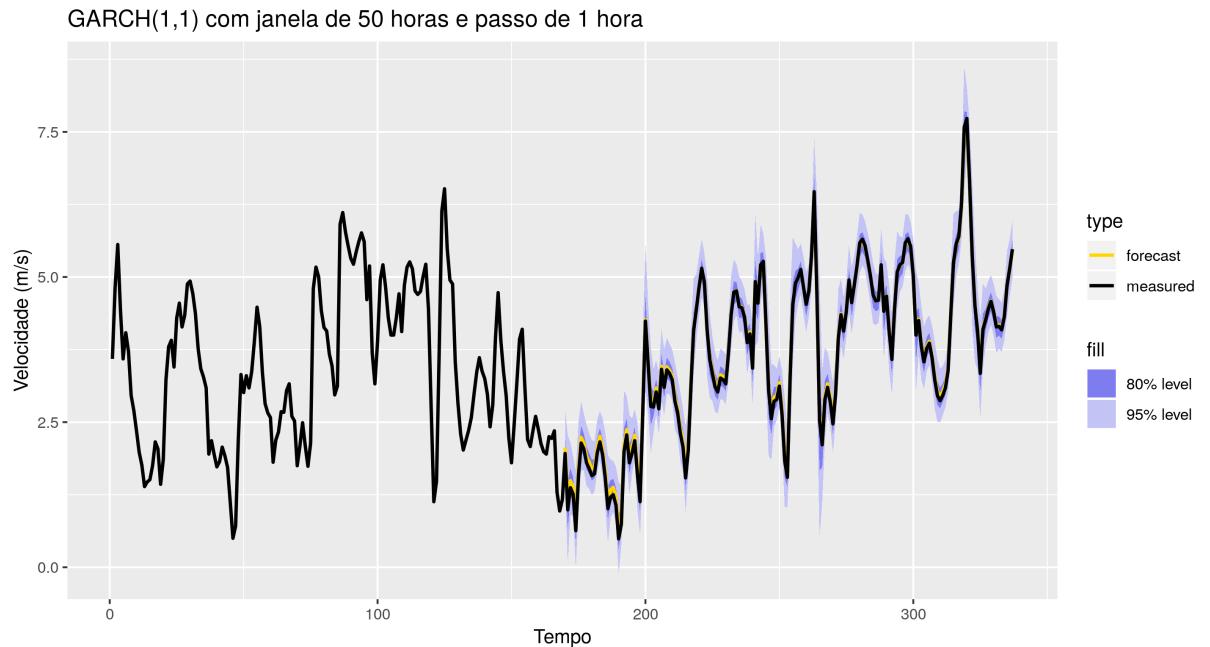


Figura 31 – GARCH

6.2 Base mensal

Embora a previsão horária seja de grande utilidade tanto para operadores de parques eólicos quanto para operadores de subestações de distribuição de energia, contratos de

Modelo	ME	RMSE	MAE	MPE	MAPE
GARCH(1,1)	-0.04246845	0.06505767	0.05117909	-2.449632	2.592525
ARIMA(1,1,1)	-0.007948941	0.2231684	0.161981	0.3817872	5.731274
ARIMA(2,1,1)	0.00824027	0.2973698	0.2285103	-0.7558966	7.622429
ARIMA(1,1,2)	0.01559968	0.2490103	0.1870849	-0.8393794	6.151953
ARIMA(2,1,3)	0.01455176	0.3189862	0.2380204	-0.8382305	7.968127

produção e entrega de energia são feitos em base mensal. Dessa forma, um acionista de um parque eólico deseja saber em qual nível de produção ele pode arriscar se comprometer em entregar em um acordo fechado um mês antes. O mesmo modelo, GARCH(1,1), se demonstra novamente adequado:

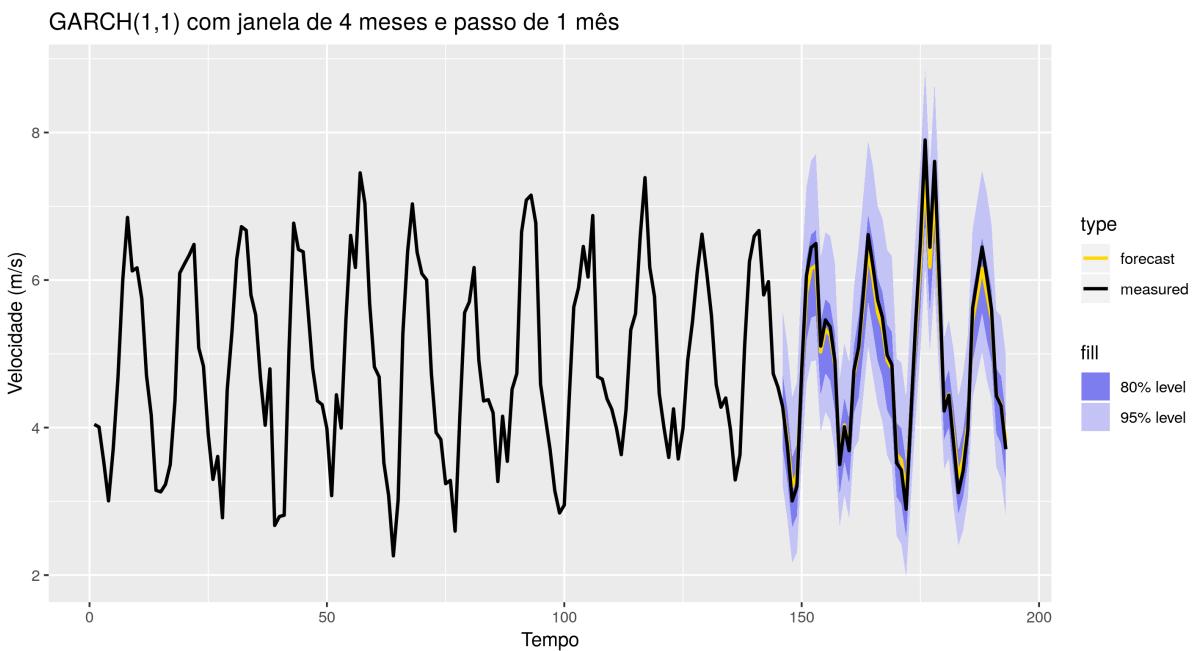


Figura 32 – GARCH

Modelo	ME	RMSE	MAE	MPE	MAPE
GARCH(1,1)	0.06669541	0.1773425	0.1461843	0.5504064	2.893245

7 Conversão em energia

De acordo com a equação 1 a potência desenvolvida por um aerogerador é função principalmente da área coberta pelas suas pás, da densidade local do ar e do cubo da velocidade do vento. Exceto pela velocidade do vento os demais parâmetros variam pouco. Dessa forma, uma vez definidos os outros parâmetros, é possível obter a potência desenvolvida por um aerogerador por meio de sua curva de potência que mapeia intervalos de velocidade em intervalos de potência. A figura abaixo é um exemplo de curva de potência genérica:

Curva de potência de uma turbina genérica

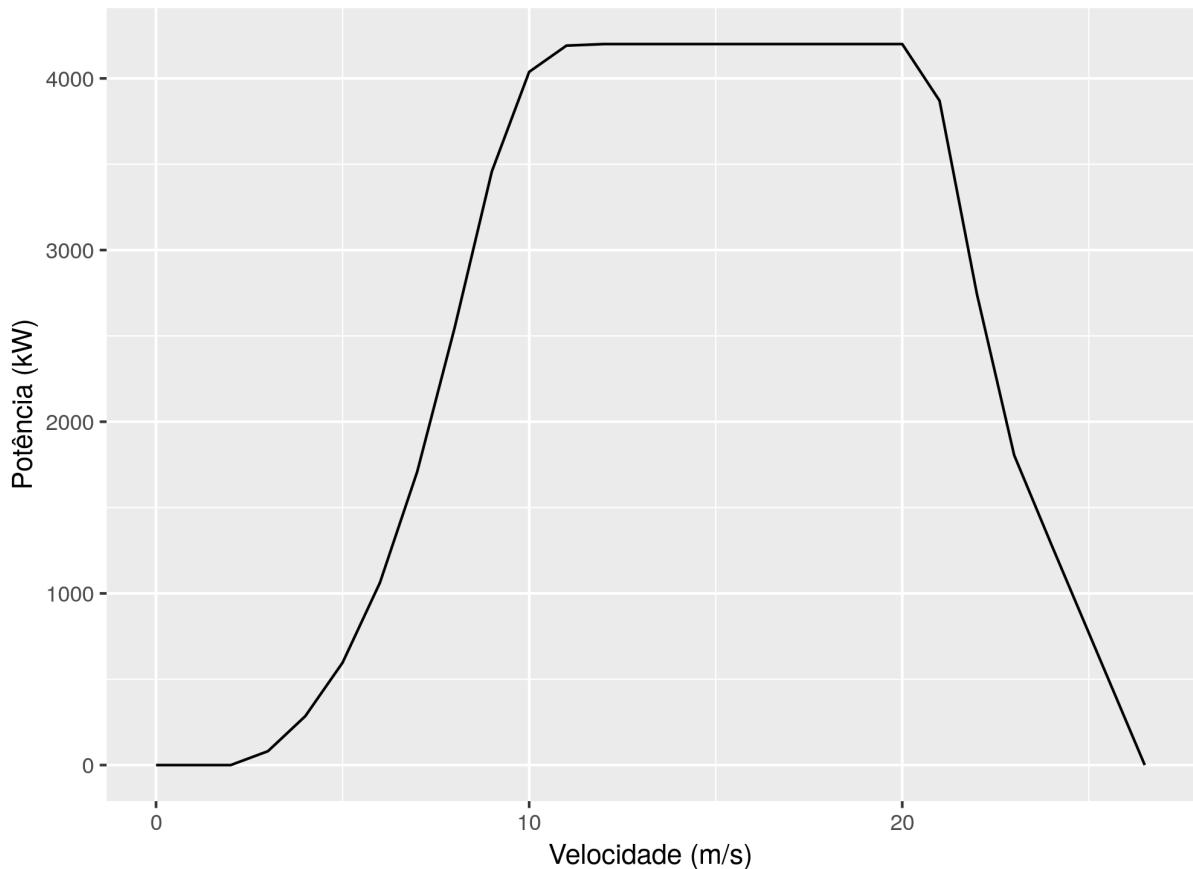


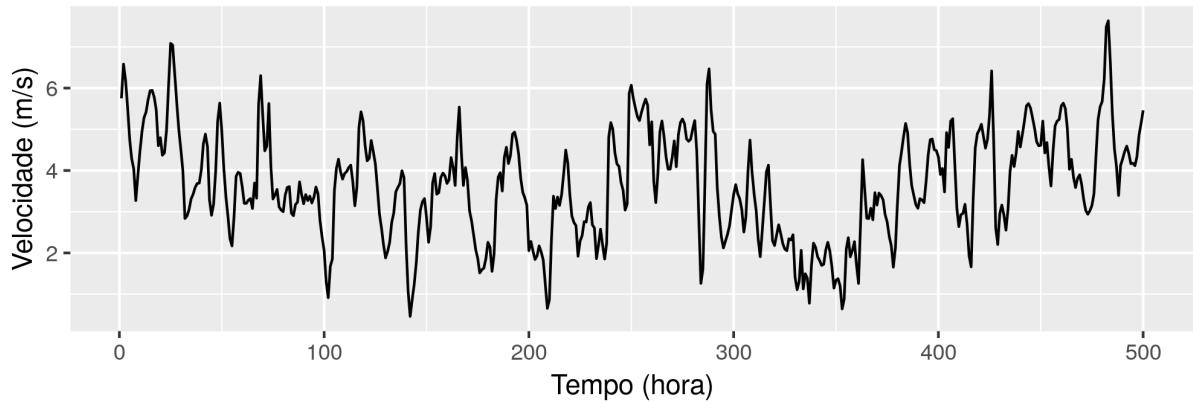
Figura 33 – Curva de potência de um aerogerador genérico.

As curvas de potência de qualquer turbina eólica possuem esse formato característico. A turbina só produz energia a partir de uma certa velocidade mínima. Para velocidades mais altas encontra uma região linear, onde a potência aumenta linearmente com a velocidade. Para uma certa velocidade chega-se num plateau de potência acima do qual um aumento de velocidade não causa variação na potência. Para velocidades muito altas é necessário

desacelerar a rotação das pás de modo a não danificar componentes do aerogerador e portanto a potência decaí. Quando a velocidade excede um valor limite é necessário desligar o aerogerador.

O mapeamento entre velocidade e potência é exibido abaixo. A série temporal de velocidade do vento corresponde àquela para a qual foi feita a previsão em 31.

Velocidade do vento em base horária



Potência desenvolvida por uma turbina em base horária

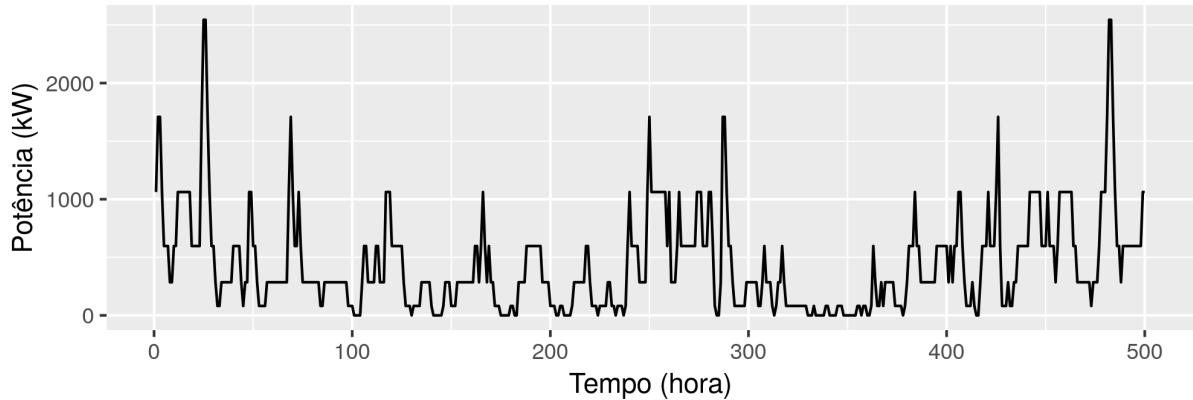


Figura 34 – Curva de potência de um aerogerador genérico.

De acordo com essa curva a energia total produzida nas 500 horas previstas seria de 217 MWh. Embora durante o período considerado a geração tenha ficado distante da potência máxima, o aerogerador produziu uma quantidade expressiva de energia. Se no mesmo período de 500 horas o aerogerador operasse no seu valor máximo de 4.5 MW, seriam produzidos 2250 MWh. Para esse período o fator de capacidade foi baixo, apenas cerca de 0.1. É comum encontrar fatores entre 0.3 e 0.5 em parques operacionais no país. É interessante perceber que caso o proprietário do parque fechasse um acordo de produção de energia considerando um fator de capacidade relativamente pessimista de 0.3 (o limite inferior da média dos parques do país) ele perderia dinheiro pois durante o período o fator de capacidade ficou abaixo de 0.1.

8 Validação com outras séries

O modelo que obteve o melhor poder preditivo foi o GARCH(1,1). No entanto, assim como todos os outros modelos discutidos, os testes de poder preditivo foram feitos apenas sob uma única série, aquela apresentada na seção 2. Há regiões em que a distribuição que melhor descreve o recurso eólico é a de weibull, em outras é a distribuição Rayleigh, beta, normal ou log-normal (DROBINSKI; COULAIS, 2012). O vento pode ainda ser unidirecional, bidirecional ter mais de uma direção predominante. Como foi exposto, a série modelo é melhor descrita por uma distribuição beta (Figura 11) e é predominantemente unidirecional (Figura 12). Dessa forma é interessante validar o modelo aplicando-o em diversas outras séries correspondentes a diferentes regiões cujo recurso eólico seja distinto.

62 regiões foram escolhidas. 7 na América Central, 1 no Jordão e as demais na América do Sul. A distribuição desigual se deve ao fato de que essas séries correspondem a estudos comerciais ao qual o autor tem acesso. Explica-se dessa forma o fato de a grande maioria das séries localizarem-se no Nordeste do Brasil. Todas as séries utilizadas, no entanto, são oriundas do banco de dados de acesso público ERA5 da ECWMF. Essas localidades são exibidas abaixo:

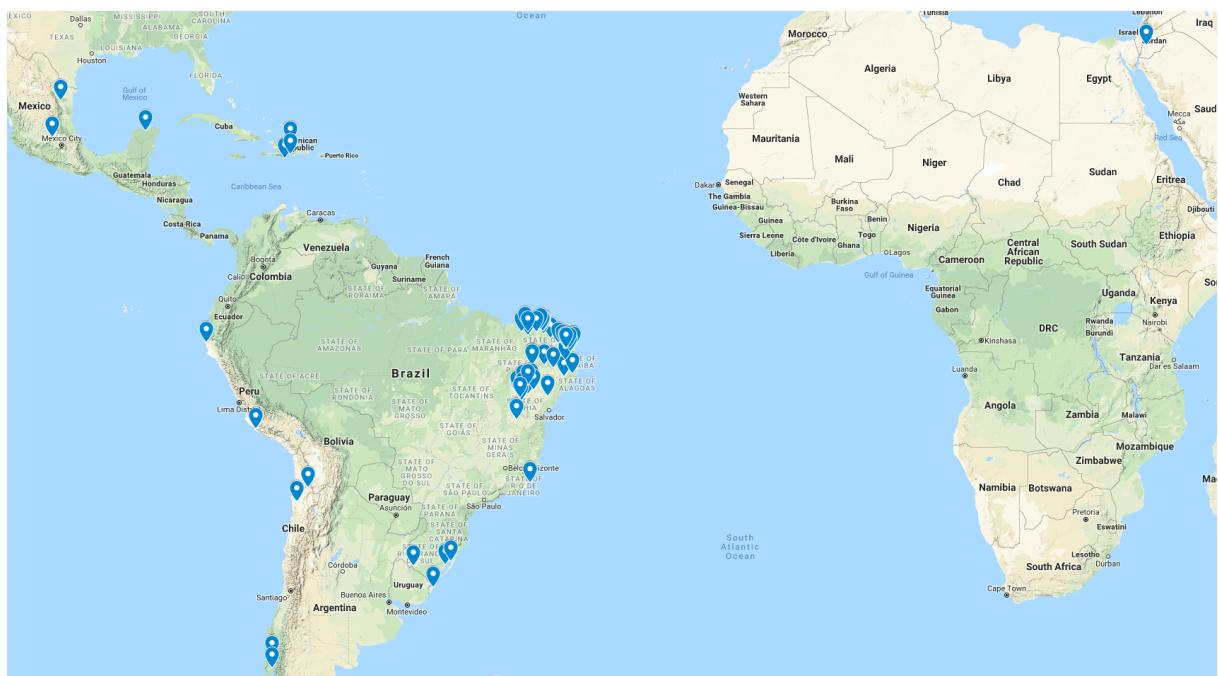


Figura 35 – Regiões correspondentes a dados de velocidade do vento do banco de dados ERA5 da ECWMF. Mapa disponível em [TDEF Map](#)

As regiões escolhidas variam entre florestadas, desérticas, no interior de continentes ou diretamente na costa ou ainda em golfos ou ilhas. O sucesso do modelo em regiões tão

distintas leva a crer que ela seja aplicável a condições bem gerais.

O modelo GARCH(1,1) foi aplicado em cada uma das séries cujas localidades correspondentes são exibidas no mapa acima e os resultados são exibidos abaixo:

Local	ME	RMSE	MAE	MPE	MAPE
NE01a	0.0	0.02	0.01	-0.0	0.17
NE05	0.01	0.02	0.01	0.02	0.17
ChapadaAraripe	0.01	0.02	0.01	0.02	0.17
Embuacá	-0.01	0.02	0.02	-0.16	0.27
Santos	-0.01	0.02	0.02	-0.16	0.27
LaGuajira	0.03	0.04	0.04	0.25	0.27
AsaBranca	0.02	0.03	0.02	0.19	0.28
Omega	0.02	0.03	0.03	0.14	0.28
Icarai	-0.01	0.02	0.02	-0.19	0.29
EsquinaDoVento	0.03	0.04	0.03	0.25	0.4

Tabela 1 – 10 melhores localidades em relação a previsão com GARCH(1,1)

Local	ME	RMSE	MAE	MPE	MAPE
Omega_RJ	0.1	0.13	0.11	1.48	2.02
Ibipeba	-0.04	0.07	0.05	-1.83	2.11
Tafila ERA5	0.05	0.11	0.08	-0.19	2.33
CapaoAlto	0.0	0.07	0.06	-1.63	2.41
Quadran	-0.05	0.07	0.05	-2.48	2.61
LaVigia	-0.01	0.06	0.05	-2.28	3.12
Ckani	0.05	0.13	0.09	-0.95	3.35
SanPedro	0.08	0.17	0.12	-0.1	3.94
Tucano	-0.05	0.15	0.13	-3.74	4.8
Llanos	0.22	0.33	0.24	3.47	12.58

Tabela 2 – 10 piores localidades em relação a previsão com GARCH(1,1)

Referências

- ACKERMANN, T. *Wind Power in Power Systems*. [S.l.]: John Wiley & Sons, Ltd, 2012. Citado na página 14.
- AN., C. Energy output estimation for small-scale wind power generators using weibull-representative wind data. *Journal of Wind Engineering and Industrial Aerodynamics*, 2003. Citado 2 vezes nas páginas 4 e 6.
- CENTRO DE PESQUISAS DE ENERGIA ELÉTRICA, CEPEL. *Atlas do Potencial Eólico Brasileiro*. Brasíli, 2001. Disponível em: <http://www.cresesb.cepel.br/publicacoes/download/atlas_eolico/atlas%20do%20potencial%20eolico%20brasileiro.pdf>. Citado na página 14.
- DROBINSKI, P.; COULAIS, C. Is the weibull distribution really suited for wind statistics modeling and wind power evaluation? 11 2012. Citado na página 56.
- ECMWF. *ERA5 Reanalysis Database*. 2017. Disponível em: <<https://www.ecmwf.int/en/forecasts/datasets/archive-datasets/reanalysis-datasets/era5>>. Acesso em: 03 nov 2018. Citado 2 vezes nas páginas 7 e 16.
- M HOOSHMAND R-A, K. A. P.; A-H, Z. Joint operation of wind farm, photovoltaic, pump-storage and energy storage devices in energy and reserve markets. *International Journal of Electrical Power & Energy Systems*, 2015. Citado 2 vezes nas páginas 4 e 6.
- MENČÍK, J. Weibull distribution. *Springer, Berlin, Heidelberg*, v. 1, n. 1, p. 71–97, 2016. Citado 2 vezes nas páginas 4 e 6.
- P, V. S. R. A review of wind speed probability distributions used in wind energy analysis: case studies in the canary islands. *Renewable Sustainable Energy Review*, 2003. Citado na página 4.