

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

ESCOLA DE ENGENHARIA

DEPARTAMENTO DE ENGENHARIA ELÉTRICA

CHRISTOPHER ANDREWS

**MODELOS ESTATÍSTICOS DE PREVISÃO DE CARGA E
GERAÇÃO RENOVÁVEL PARA OPERAÇÕES EM
SISTEMAS DE POTÊNCIA**

Porto Alegre - RS

2019

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

ESCOLA DE ENGENHARIA

DEPARTAMENTO DE ENGENHARIA ELÉTRICA

**MODELOS ESTATÍSTICOS DE PREVISÃO DE CARGA E
GERAÇÃO RENOVÁVEL PARA OPERAÇÕES EM
SISTEMAS DE POTÊNCIA**

Projeto de Diplomação apresentado ao Departamento de Engenharia Elétrica da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para Graduação em Engenharia Elétrica.

ORIENTADOR: M.Sc. Davood Raof-Sheibani

EXAMINADOR: Prof. Dr. Wolfram H. Wellssow

SUPERVISOR: Prof. Dr. Paulo Roberto Eckert

Porto Alegre

CIP - Catalogação na Publicação

Andrews, Christopher
Modelos estatísticos de previsão de carga e geração renovável para operações no sistema de potência / Christopher Andrews. -- 2019.
102 f.
Orientador: Davood Raof-Sheibani.

Trabalho de conclusão de curso (Graduação) -- Universidade Federal do Rio Grande do Sul, Escola de Engenharia, Curso de Engenharia Elétrica, Porto Alegre, BR-RS, 2019.

1. Energia renovável. 2. Previsão estatística. I. Raof-Sheibani, Davood, orient. II. Título.

RESUMO

O objetivo deste trabalho é a criação de ferramentas de previsão baseado em modelos estatísticos para operações em sistema de potência. Os horizontes de previsão dos modelos serão de oito horas no futuro e um dia no futuro. As variáveis de previsão são: geração fotovoltaica, geração eólica, carga e carga vertical no nível de transmissão. Para proceder com o trabalho inicialmente é feita a coleta das variáveis de previsão juntamente com um conjunto de dados meteorológicos auxiliares. O pré-processamento é necessário para garantir que os dados respeitem as restrições impostas pelos modelos estatísticos de previsão. Os modelos de previsão são baseados em métodos estatísticos conhecidos e confiáveis, usados em análise de séries temporais. No final, o melhor modelo para cada conjunto de dados de previsão é encontrado e seu desempenho é medido. Os resultados da previsão de carga e da carga vertical foram muito satisfatórios, exibindo um erro percentual muito baixo, mantendo um intervalo de erros globalmente constante ao longo da série temporal. Quanto à previsão de geração de energia fotovoltaica e eólica, os resultados foram aceitáveis para algumas estimativas grosseiras. O desempenho da previsão pode ser melhorado usando modelos não-lineares ou modelos que possam lidar com a volatilidade.

Palavras-chaves: Fontes renováveis. Previsão de variáveis da rede de transmissão. Estabilidade do sistema. Previsão estatística.

Abstract

The objective of this thesis is to create forecasting tools for day-ahead or intraday power system operation, forecasting photovoltaic (PV) generation, wind generation, load, and vertical load at the transmission level. To find the best models, it is necessary to collect the forecasting along with the predictor datasets. Preprocessing is also necessary to ensure the datasets respected some qualities to be fed into the model. The forecasting models are based on well-known and reliable statistical methods used for time series analysis. In the end, the best model for each forecasting dataset is found and its performance is measured. The results of forecasting load and vertical load are satisfying, showing a very low percentage error, maintaining an overall constant range of error. As to the forecast of PV and wind generation, the results can be acceptable for some raw estimation, but the performance of the forecast can be improved by using nonlinear models or models that can address volatility.

Keywords: Renewable energy sources. Prediction of electrical Grid variables. Grid stability. Statistical forecasting tools.

Lista de Figuras

Figura 1: Principais políticas relacionadas a energia e eventos históricos nos últimos 30 anos	15
Figura 2: Mudanças de representatividade das fontes de produção de energia de 1990 a 2018 com base em (Energiebilanzen).....	16
Figura 3: Série temporal descrita por um processo AR (2).....	28
Figura 4: Distribuição de erro em torno da previsão, mostrando também o intervalo de confiança de 95% em torno da previsão.....	29
Figura 5: Série temporal descrita por um processo MA (2).....	30
Figura 6: Série temporal descrita por um processo ARMA (2,2).....	31
Figura 7: Série temporal descrita por um processo ARIMA (2,1,1).....	33
Figura 8: Série temporal que representam um processo SARIMA (2,1,1) x (1,1,0) ₁₀₀	35
Figura 9: FAC e FACP de um processo de AR, onde o FAC diminui gradualmente e o FACP corta depois de p defasagens (BROCKWELL, 1996).	37
Figura 10: Modelo <i>underfitted</i> , modelo <i>overfitted</i> e modelo otimizado respectivamente.....	41
Figura 11: Área de controle da Amprion em roxo.....	45
Figura 12: Gráfico de distribuição e boxplot do conjunto de dados de carga. .	47
Figura 13: Séries temporais fotovoltaicas com tendência	49
Figura 14: Sazonalidade de séries temporais fotovoltaicas	50

Figura 15: Multiplicação de tendência com sazonalidade sobreposta à série temporal original.....	50
Figura 16: Séries temporais fotovoltaicas sem tendência e sem sazonalidade anual.....	51
Figura 17: Série temporal fotovoltaica de sazonalidade diária	52
Figura 18: Gráfico da função de auto correlação da série temporal resultante	53
Figura 19: Séries temporais de geração de energia eólica com sua respectiva tendência.....	54
Figura 20: Similaridades entre os dados de velocidade média do vento exógeno e a série temporal de geração eólica resultante	54
Figura 21: Gráfico de dispersão mostrando o grau de correlação entre a média da velocidade de ventos na região da Amprion e a energia eólica gerada na mesma região.	55
Figura 22: Tendência anual do conjunto de dados de carga sobreposta à série temporal da carga	56
Figura 23:A sazonalidade anual do conjunto de dados de carga se sobrepõe à série temporal da carga.....	57
Figura 24: Comparação do conjunto de dados de carga quando sua sazonalidade anual foi removida com o padrão de sazonalidade da carga semanal.	58
Figura 25: Série temporal residual da carga	58
Figura 26: FAC da série temporal resultante da carga.....	59
Figura 27: Séries temporais de carga vertical com a linha de tendência anual	60

Figura 28: A sazonalidade anual da série temporal de carga vertical sobreposta à série temporal da carga vertical.	61
Figura 29: Séries temporais de carga restante	61
Figura 30: FAC dos resíduos da série temporal da carga vertical.....	62
Figura 31: Mapa de calor mostrando a correlação dos valores previstos em relação à média de cada preditor	63
Figura 32: Fluxograma de código SARIMAX.....	67
Figura 33: Parâmetros usados no processo de Grid Search em uma forma de laço for.....	70
Figura 34: Trecho do shell script enviado à Elwetritsch	72
Figura 35: <i>Flowchart</i> da configuração de <i>walk forward</i> para achar o melhor modelo.....	74
Figura 36: 200 previsões de carga de 8 horas	83
Figura 37: Crescimento do erro RMSE para o conjunto de dados de carga com o aumento do horizonte de previsão até no máximo 8 horas	83
Figura 38: Crescimento do erro percentual médio de carga com o aumento do horizonte de previsão até no máximo 8 horas	84
Figura 39: 200 previsões de carga de 24 horas para o futuro	85
Figura 40: Crescimento do erro RMSE para o conjunto de dados de carga com o aumento do horizonte de previsão até no máximo 24 horas	85
Figura 42: 8 horas de previsão da carga vertical	87
Figura 43: Crescimento do erro de RMSE para o conjunto de dados de carga vertical com o aumento do horizonte de previsão até no máximo 8 horas.....	87

Figura 44: Crescimento do erro percentual médio da carga vertical com o aumento do horizonte de previsão até no máximo 8 horas	88
Figura 45: Multiplas previsões de 24 horas da carga vertical junto com a carga vertical original	89
Figura 46: Crescimento do erro RMSE para o conjunto de dados de carga vertical com o aumento do horizonte de previsão até no máximo 24 horas a partir do tempo real	89
Figura 47: Crescimento do erro percentual médio da carga vertical com o aumento do horizonte de previsão até no máximo 24 horas a partir do tempo real ..	90
Figura 48: 100 múltiplas previsões de 8 horas para a geração fotovoltaica com o conjunto original	91
Figura 49: Crescimento do erro RMSE para a geração fotovoltaica com o aumento do horizonte de previsão até no máximo 8 horas a partir do tempo real	91
Figura 50: Crescimento do erro médio percentual para a geração fotovoltaica com o aumento do horizonte de previsão até no máximo 8 horas a partir do tempo real	92
Figura 51: Múltiplas previsões de 24 horas para o futuro para a geração fotovoltaica	93
Figura 52: Crescimento do erro RMSE das previsões de geração fotovoltaica com o aumento do horizonte de previsão de até 24 horas	93
Figura 53: Crescimento do erro percentual médio das previsões de geração fotovoltaica com o aumento do horizonte de previsão de até 24 horas	94
Figura 54: 200 múltiplas previsões de geração eólica	95
Figura 55: Crescimento do erro RMSE das previsões de geração eólica com o aumento do horizonte de previsão de até 8 horas	96

Figura 56: Crescimento do erro percentual médio das previsões de geração eólica com o aumento do horizonte de previsão de até 8 horas	96
Figura 57: 200 múltiplas previsões de 24 horas de geração eólica.....	97
Figura 58: Crescimento do erro RMSE para previsões de 24 horas de geração eólica.....	98
Figura 59: Crescimento do erro percentual médio para previsões de 24 horas de geração eólica	98

Lista de Tabelas

Tabela 1: Meta estabelecida na Lei EEG de 2014 com base em (IFEU)	17
Tabela 2: Características de FAC e FACP de séries temporais estacionárias	38
Tabela 3: Variáveis usadas nas previsões.	44
Tabela 4: Estrutura das variáveis preditoras.	46
Tabela 5: Bibliotecas relevantes usadas	66
Tabela 6: Opções de parâmetros para <i>Grid Search</i> da previsão da geração fotovoltaica	69
Tabela 7: Parâmetros de entrada Geração eólica, carga e de carga vertical..	71
Tabela 8: Coleção de melhores modelos para conjuntos de dados de carga vertical	75
Tabela 9: Resultado da validação cruzada retorna melhor modelo global para carga vertical	75
Tabela 10: Modelos de previsão de carga retornados da etapa <i>Grid Search</i> com menores erros	76
Tabela 11: Resultados da validação cruzada para o conjunto de dados de carga	77
Tabela 12: Melhores modelos para previsão fotovoltaica obtidos a partir do processo de <i>Grid Search</i>	78
Tabela 13: Resultados de validação cruzada para geração de PV	78
Tabela 14: Coleção dos modelos de previsão de geração eólica oriundo do processo de <i>Grid Search</i>	79

Tabela 15: Resultados da validação cruzada dos modelos de geração eólica	80
Tabela 16: Resumo dos resultados obtidos no processo de validação cruzada.	81

Sumário

1	Introdução	14
1.1	Panorama da geração elétrica na Alemanha	14
1.2	Funcionamento do Mercado de energia	17
1.3	Objetivos	19
1.4	Estrutura do trabalho	20
2	Análise e previsão de séries temporais.....	22
2.1	Introdução	22
2.2	Análise de séries temporais	22
2.3	Referencial teórico de métodos estatísticos para análise de séries temporais	23
2.3.1	Características de modelos estatísticos de previsões	23
2.3.2	Série temporal estacionária	23
2.3.3	Componentes de uma série temporal.....	24
2.4	Modelos de previsão estatística	27
2.4.1	Modelos auto regressivos.....	27
2.4.2	Modelos de média móvel.....	29
2.4.3	Modelos auto regressivos de médias móveis	30
2.4.4	Modelo auto regressivo integrado de médias móveis.....	31
2.4.5	Modelo auto regressivo integrado de média móvel sazonal.....	33
2.5	Determinação da ordem dos modelos.....	36

2.5.1	Função de auto correlação e função de auto correlação parcial ...	36
2.5.2	<i>Grid Search</i>	39
2.6	Medição do desempenho da previsão.....	39
2.6.1	Definição de erro de previsão.....	39
2.6.2	Validação cruzada	40
2.7	Conclusão	41
3	Conjuntos de dados de previsão e conjunto de dados auxiliares	43
3.1	Introdução	43
3.2	Variáveis de previsão	43
3.2.1	Variáveis auxiliares.....	45
3.3	Preparação de variáveis para respeitar os critérios dos modelos	47
3.4	Conjunto de dados de geração fotovoltaica	48
3.5	Conjunto de dados de geração eólica	53
3.6	Série temporal da carga	55
3.7	Série temporal de carga vertical.....	59
3.8	Seleção de dados exógenos como preditores.....	62
3.9	Conclusão	63
4	Desenvolvimento de modelos	65
4.1	Introdução	65
4.2	Implementação de métodos estatísticos para previsão de séries temporais	65
4.3	Modelo de previsão de carga vertical	74

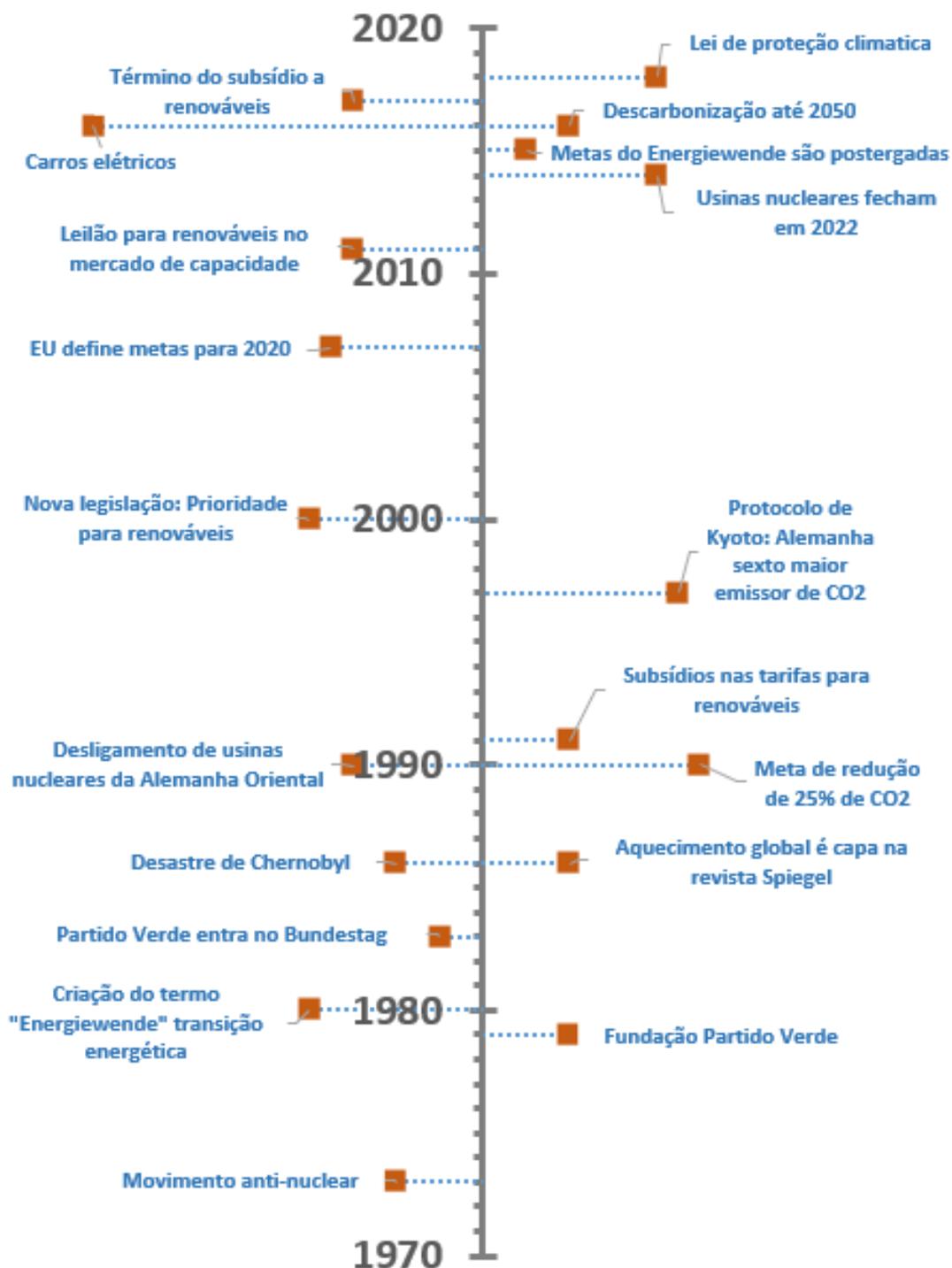
4.4	Modelo de previsão de carga	76
4.5	Modelo de previsão de geração fotovoltaica	78
4.6	Modelo de previsão de geração eólica	79
4.7	Conclusão	80
5	Resultados de previsão.....	82
5.1	Introdução	82
5.2	Resultados da previsão de carga	82
5.2.1	Múltiplas previsões de 8 horas	82
5.2.2	Múltiplas previsões de 24 horas pro futuro	84
5.3	Resultados da previsão de carga vertical	86
5.3.1	Múltiplas previsões de 8 horas	86
5.3.1	Múltiplas previsões de 24 horas pro futuro	88
5.4	Resultados da previsão de geração fotovoltaica	90
5.4.1	Múltiplas previsões de 8 horas pro futuro	90
5.4.1	Múltiplas previsões de 24 horas pro futuro	92
5.5	Resultados da previsão de geração de energia eólica	95
5.5.1	Múltiplas previsões de 8 horas	95
5.5.1	Múltiplas previsões de 24 horas pro futuro	96
6	Conclusão	99

1 Introdução

1.1 Panorama da geração elétrica na Alemanha

Desde a década de 1970, a Alemanha vem investindo em uma economia ambientalmente responsável. Esta decisão exigiu apoio legal da classe dominante para estabelecer as diretrizes para os investimentos necessários. No ano de 1995, 3,2% da produção bruta de energia na Alemanha era oriunda de fontes renováveis, já no ano de 2018, 40,3% da produção bruta de energia tem origem em fontes renováveis (BRUNNO BURGER, 2019) Na forma de linha do tempo, a Figura 1 ilustra os principais acontecimentos que moldaram a produção de energia na Alemanha (WETTENGEL, 2017).

Figura 1: Principais políticas relacionadas a energia e eventos históricos nos últimos 30 anos

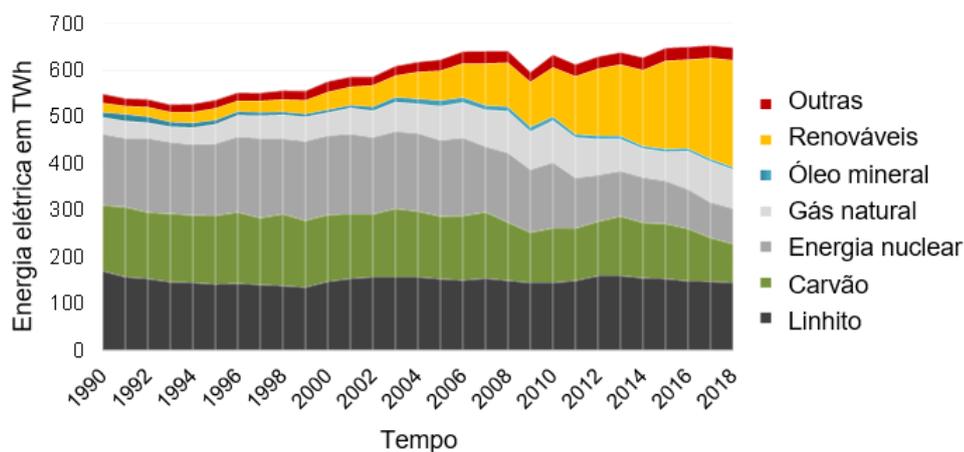


Fonte: adaptado de

A legislação do ano 1990 garantia a conexão à rede para qualquer fornecedor de energia renovável, seja qual for a disponibilidade de fornecimento, garantia

também o despacho preferencial de renováveis e um pagamento feito a residências ou empresas para gerar sua própria eletricidade, a chamadas tarifa *Feed-in* (FIT). A tarifa *Feed-in* é um mecanismo político concebido para acelerar o investimento em tecnologias de energias renováveis, garantindo o retorno do investimento através de um pagamento de uma taxa para o investidor. A fonte de financiamento da tarifa *Feed-in* era fornecida pela sobretaxa dos consumidores, exceto as empresas de grande consumo energético, de forma a garantir a concorrência global (AGORA ENERGIEWENDE, 2016). No ano de 2000 foi lançada a lei chamada *Erneuerbare-Energien-Gesetz*(EEG), sendo a lei mais conhecida na orientação da instalação de fontes de energia renováveis na Alemanha. A Figura 2, mostra a mudança de representatividade das fontes de produção de energia de 1990 a 2018.

Figura 2: Mudanças de representatividade das fontes de produção de energia de 1990 a 2018 com base em (Energiebilanzen).



Fonte: Energiebilanzen (2019)

Em 2014, a lei *Erneuerbare-Energien-Gesetz*(EEG) foi revisada. A característica mais importante da lei revisada foi o estabelecimento de um sistema baseado em leilões para a maioria das tecnologias de geração e o estabelecimento

de metas para a geração de energias renováveis. Tais metas são acompanhadas por relatórios de progresso anuais. Os alvos definidos são mostrados na Tabela 1(IFEU).

Tabela 1: Meta estabelecida na Lei EEG de 2014 com base em (IFEU)

Ano	Parcela de energias renováveis
2025	40-45%
2035	55-60%
2050	>80%

Fonte: Institut für Energie und Umweltforschung Heidelberg

1.2 Funcionamento do Mercado de energia

Com o aumento de fontes de energia voláteis como as fontes de energia renováveis, o problema da incerteza quanto a disponibilidade de fornecimento se agrava. Esta questão acaba por afetar a operação do sistema de transmissão de potência assim como a estrutura do mercado de energia elétrica. Para lidar com a incerteza de produção de energia, o processo de previsão vem ganhando importância no planejamento estratégico operacional.

As previsões são muito utilizadas em ativação de reservas para balanceamento, despacho de geração prioritário, balanceamento em tempo real e planejamento operacional comercial no mercado de eletricidade (RICARDO BESSA, 2014). O conjunto de previsão das gerações de energia e das cargas contribui para a antecipação do estado mais provável da rede de transmissão. Desta forma cria-se um ponto de equilíbrio de custo-benefício entre a carga e geração, tanto no planejamento intradiário como no planejamento do dia seguinte, permitindo assim a otimização do uso dos combustíveis sem abrir mão da robustez geral do sistema de transmissão.

O mercado de energia é dividido em três categorias definidas de acordo com o quão próximo o contrato é firmado em relação ao momento atual do consumo de energia *Forward market*

antecedência. O mercado do dia seguinte *Day ahead market* é onde as transações comerciais são realizadas de acordo com pacotes de energia de uma hora, com o fechamento das negociações às 12h00 do dia anterior ao consumo. O mercado intradiário *intraday market*, o qual negocia pacotes de energia de 15 minutos, começando às 3:00 da tarde do dia anterior ao consumo e terminando 45 minutos antes da entrega da energia. O *Forward market* e o *day ahead market* fornecem uma estimativa muito mais grosseira de geração e consumo. É então a tarefa do *intraday market* preencher as lacunas entre consumo e suprimento, positiva ou negativamente. (KÜHNERT, 2016).

As previsões de disponibilidade de geração e demanda de energia são usadas para determinar o preço do custo marginal de produção das diversas fontes para estabelecer a ordem de mérito de quais geradores devem ser acionados. A precisão da previsão *day ahead market* é o que determina o custo da eletricidade no mercado. Desta forma as previsões intradiárias são usadas para atualizar os preços e as condições de balanceamento para uma situação mais próxima do consumo real. No mercado intradiário, o intervalo de confiança da previsão é menor que o do dia seguinte e os demais erros de previsão são tratados com o uso do mercado de energia de controle *control energy market*. (MÖLLER, 2010)

A colocação das unidades geradoras é ditada por uma ordem ascendente de custos marginais, chamada de ordem de mérito. Por lei, as fontes de energia renováveis são colocadas em primeiro lugar na ordem de mérito, seguidas por usinas nucleares, carvão linhito e usinas de carvão. Ao somar toda a energia necessária, o custo da eletricidade é definido pela união de toda energia demandada no *day ahead market*.

O mercado de energia e o mercado de reserva de controle (*control reserve market*) garantem o equilíbrio entre a produção e o consumo de eletricidade. Os operadores de sistemas de transmissão (OST) são obrigados a vender toda a eletricidade fornecida a partir de fontes renováveis no *European Power Exchange* (EPEX) (BRUNNER, 2014) - (CHRISTIAN GROWITSCH, 2011). Devido ao mecanismo da tarifa *Feed-in*, o preço pago pela produção de energia renovável é

maior do que o preço de venda da energia (ou seja, o preço de compensação do mercado), que segue a curva de oferta e demanda. Isto cria uma perda de receita para os operadores do sistema e estas perdas monetárias são compensadas pelo *EEG Umlage* (AGORA ENERGIEWENDE, 2016) uma tarifa embutida na conta de todos os consumidores. Como não seria justo que as regiões com elevada ocorrência de renováveis (como ocorre no norte da Alemanha, onde há muita geração *offshore*) pagassem uma quantia mais elevada, este montante é determinado agrupando todo o fornecimento de energia renovável na Alemanha e dividindo-os por todo o país.

Na operação em tempo real, desequilíbrios podem ocorrer por meio de imprevistos, como falhas em usinas de energia, falhas na rede de transmissão, ou erros de previsão de cargas e de geração. Para manter a estabilidade da rede, as unidades de despacho rápido precisam estar prontas para fornecer mais energia em casos de escassez ou reduzir a produção de energia em caso desta exceder a demanda. Esse balanço energético é dividido em três estágios de tempo diferentes: a reserva de contenção de frequência, chamada *frequency containment reserve* (FCR), a reserva de restauração de frequência, chamada *frequency restoration reserve* (FRR) e a reserva de restauração, chamada *restoration reserve* (RR). (MÖLLER, 2010). As usinas de energia que fazem parte da reserva de controle necessitam boas características de aceleração (*ramping up*) e desaceleração (*ramping down*), que são normalmente encontradas em turbinas a gás, geradores a diesel e usinas hidrelétricas no modelo *Pumped Hydro*. No mercado de energia de controle, os fornecedores são pagos pela manutenção dessa capacidade, por estarem em prontidão para o fornecimento. Os preços para manter esta capacidade e os preços para regular a rede são diferentes e muito superiores aos praticados nos mercados diário e intradiário. A previsão de curto prazo é usada para gerenciamento de demanda e despacho de geração, bem como para manter as características de estabilidade da rede. (MÖLLER, 2010).

1.3 Objetivos

A tendência mundial de aumentar a participação de fontes renováveis na produção de energia impõe uma maior volatilidade das características de estabilidade

da rede. Para garantir o conhecimento sobre um estado futuro das características da rede e prever assim possíveis problemas, é necessário dispor de ferramentas para analisar a probabilidade de um determinado estado da rede de transmissão. A ideia do presente trabalho é construir uma ferramenta de previsão para a operação do sistema de energia com base em dados históricos. Serão usados na previsão dados de geração fotovoltaica, geração eólica, carga e carga vertical da região da empresa transmissora Amprion. Os conjuntos de dados de previsão foram escolhidos devido à sua volatilidade e capacidade de alterar rapidamente os parâmetros da rede de transmissão.

Este trabalho também engloba o uso de dados meteorológicos como variáveis auxiliares externas para calcular com maior precisão os coeficientes dos modelos matemáticos de previsões. A aquisição dos dados de previsão e dos conjuntos de dados auxiliares é mostrado, assim como a forma que as variáveis auxiliares se relacionam com as variáveis de previsão. Todos os modelos aqui apresentados requerem um conjunto de características das séries temporais para aplicação aos modelos, tais características são posteriormente discutidas. Pré-processamento é necessário para garantir que as séries temporais respeitem esses requisitos.

1.4 Estrutura do trabalho

A estrutura deste trabalho é apresentada com uma breve sinopse dos capítulos:

O Capítulo 2 contextualiza a análise e previsão de séries temporais. Uma revisão de métodos estatísticos para análise de séries temporais com uma coleção de métodos estatísticos de previsão é apresentada. Afirma-se que as séries temporais não são totalmente previsíveis e que a sua importância se deve às suas ocorrências naturais em muitas áreas de aplicação. A relevância do pré-processamento é justificada pela necessidade de alcançar condições de estacionariedade.

O Capítulo 3 aborda as variáveis de previsão e como essas variáveis foram obtidas. As variáveis de previsão escolhidas estão relacionadas à estabilidade da

rede. É apresentado também de que forma as variáveis auxiliares podem ser relacionados às variáveis de previsão como um preditor.

O Capítulo 4 constitui a principal parte de pesquisa deste trabalho, onde os capítulos anteriores são justificados pela implementação dos modelos de previsão. A implementação foi baseada na linguagem de programação *Python*. No final deste capítulo, o melhor modelo para cada conjunto de dados é apresentado.

O Capítulo 5 apresenta os resultados da previsão para todos os conjuntos de dados de previsão usando o melhor modelo. O trabalho computacional foi feito em *Elwetritsch*, o computador de alto desempenho da região da Renânia-Palatinado.

O Capítulo 6 conclui com um resumo geral dos resultados da pesquisa. Os modelos de carga e carga vertical tiveram um desempenho muito bom, enquanto a geração fotovoltaica e a geração eólica podem ter um melhor desempenho com um modelo não linear ou um modelo que aborda volatilidade.

2 Análise e previsão de séries temporais

2.1 Introdução

Séries temporais de um determinado processo descrevem a evolução temporal das variáveis deste processo, onde o tempo é tão importante quanto a magnitude da variável indexada a ele. A importância de uma série temporal é justificada pela sua ocorrência natural em diversas áreas de aplicação, como no mercado de ações, medicina, engenharia, física e outros. (BROCKWELL, 1996). A análise de séries temporais trata do estudo das características de uma dada série temporal e de como se pode obter conclusões sobre seu comportamento futuro. As conclusões matemáticas sobre como uma série temporal evolui no tempo é usada para prever valores futuros da série temporal. Devido aos avanços no armazenamento de dados, processamento de informações e *data science*, mais dados podem ser coletados com uma frequência de amostragem maior, afetando a qualidade da previsão. Este capítulo discute as características das séries temporais, como elas podem ser desmembradas através de múltiplos processos, os métodos estatísticos de previsão e os requisitos dos modelos. É apresentado aqui ferramentas matemáticas auxiliares que podem ser usadas para encontrar os parâmetros dos modelos.

2.2 Análise de séries temporais

A análise de séries temporais compreende um conjunto de métodos, modelos e técnicas para coletar e estudar as observações de uma série temporal, a fim de desenvolver um modelo matemático adequado que represente o conjunto de dados. A previsão final é feita então com o modelo mais promissor, ou seja, o que apresentou menores erros ao longo das previsões. A previsão em si é também uma série temporal, que não pode ser prevista com total certeza devido à sua natureza estocástica. Assim, o valor previsto segue um modelo probabilístico, descrevendo a distribuição da variável aleatória em um determinado momento desta forma, sendo a distribuição de uma variável dada uma distribuição de tempo.

2.3 Referencial teórico de métodos estatísticos para análise de séries temporais

2.3.1 Características de modelos estatísticos de previsões

A suposição básica para implementar métodos estatísticos de previsões é que a série temporal é considerada uma combinação linear de valores passados e erros. A popularidade desses métodos de previsão se dá devido à sua flexibilidade de representar diversas variedades de séries temporais com simplicidade. Por outro lado, a principal limitação desses modelos é a forma linear pré-presumida das séries temporais associadas. Quando há características de baixa volatilidade na série temporal, os métodos estatísticos são capazes de alcançar uma estimativa muito precisa. (ZHONGYANG ZHAO, 2018).

2.3.2 Série temporal estacionária

Diz-se que uma série temporal é estacionária quando a média, a variância e a covariância ao longo de um período são independentes do tempo. A série temporal é estacionária se respeitar os seguintes critérios mostrados na Equação 1, Equação 2 e Equação 3.

$$\text{para todo} \quad (1)$$

Variância

Série temporal

Índice temporal

$$\text{para todo} \quad (2)$$

Valor médio

Média da série temporal

$$(3)$$

para todo

Covariância

Sazonalidade

Série temporal defasada por s unidades no tempo

Covariância com respeito a sazonalidade

Uma série temporal que respeite os preceitos de estacionariedade é fácil de prever; simplesmente prevê que as suas propriedades estatísticas serão as mesmas no futuro como foram no passado (BROCKWELL, 1996). Desta forma os métodos estatísticos são recomendados para previsão de séries temporais estacionárias e séries temporais não estacionárias que podem se tornar métodos corretos, que serão explicados mais adiante nesta seção. (BASHIRAHAMAD FARDIN MOMIN, 2017). O procedimento geral para previsão pode ser resumido na construção de modelos apropriados, definindo os parâmetros do modelo e ajustando o modelo (ou seja, o ato de calcular os coeficientes do modelo). O modelo pode então ser usado para prever o comportamento futuro das séries temporais e sua precisão pode ser avaliada, comparando com os valores ocorridos na realidade.

2.3.3 Componentes de uma série temporal

Uma série temporal pode ser decomposta em quatro componentes principais. Esses componentes são tendência (T_t), componente cíclico (C_t), componente sazonal (S_t) e componentes residuais (R_t). Nesse sentido, é possível criar modelos usando esses componentes combinando-os de forma aditiva, como na Equação 4.

(4)

Componente da tendência de uma série temporal

Componente cíclico de uma série temporal

Componente sazonal de uma série temporal

Componente residual de uma série temporal

Ou na forma multiplicativa, tal qual mostrado na Equação 1.

(5)

O componente de tendência é um padrão de longo prazo que reflete um aumento ou um decréscimo de magnitude ao longo do tempo. O componente cíclico é um padrão em uma série temporal que se repete em ciclos, porém não segue um intervalo de tempo regular, como por exemplo, o ciclo de negócios de prosperidade, recessão, depressão e expansão visto em séries econômicas. Os componentes sazonais em uma série temporal são flutuações de magnitude seguindo um intervalo de tempo regularmente espaçado. Em geral, as manifestações sazonais são criadas por condições climáticas e/ou relacionadas aos hábitos das pessoas. O componente residual representa outras flutuações aleatórias ou irregulares no sistema causadas por efeitos imprevisíveis (CHAIR OF STATISTICS, UNIVERSITY OF WÜRZBURG, 2012).

Nos casos em que os critérios de estacionariedade são violados, os componentes determinísticos de uma série temporal podem ser estimados para tornar a parte residual estacionária, criando, assim, modelos determinísticos diferentes para cada um desses componentes. É possível a criação de um modelo e previsão de uma série temporal, contanto que a série temporal não seja composta apenas por ruído branco. Formalmente, o ruído branco é um processo aleatório cuja distribuição estatística dos valores é independente e idêntica à distribuição dos valores anteriores. Tais características são mostradas na Equação 6, na Equação 7 e na Equação 2.

(6)

Covariância do ruído branco com respeito a k
defasagens de ruído branco

Ruído branco

Valor médio da série temporal de ruído branco

Defasagens

(7)

Valor esperado da média do ruído branco

(8)

Idealmente, a previsão subtraída da série temporal original deve mostrar um comportamento de ruído branco, caso contrário, ainda há informações ocultas nos resíduos que não foram usados para construir o modelo de previsão (BASHIRAHAMAD FARDIN MOMIN, 2017).

2.4 Modelos de previsão estatística

2.4.1 Modelos auto regressivos

O modelo auto regressivo (AR) afirma que a observação no tempo t é bem descrita por uma combinação linear das observações anteriores adicionadas por um termo de erro. O modelo auto regressivo, como o nome explica, é uma regressão realizada usando valores da própria série temporal, onde valores anteriores multiplicados por coeficientes são usados para prever um valor futuro. É possível enxergar tais coeficientes como fatores de ponderação da representatividade das observações passadas. A ordem de um modelo AR é o número de valores anteriores conhecidos como p na série temporal usada para prever o valor no futuro. O modelo geral AR (p) é descrito na Equação 9 e Equação 10. [13],

(9)

Constante

Coeficientes dos termos defasados da série temporal

Erro baseado em uma distribuição gaussiana com media zero e desvio padrão

Valores passados da série temporal

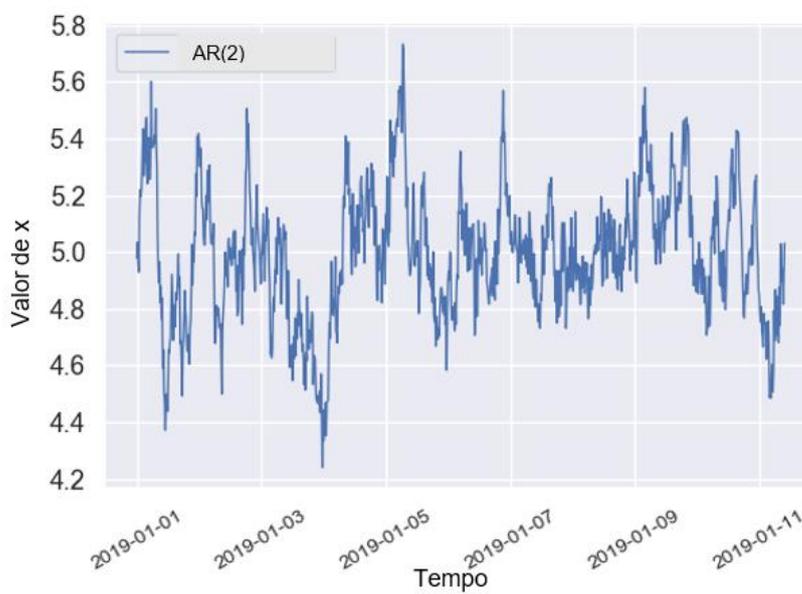
(10)

Um modelo de segunda ordem AR (2) é descrito na Equação 11 abaixo:

(11)

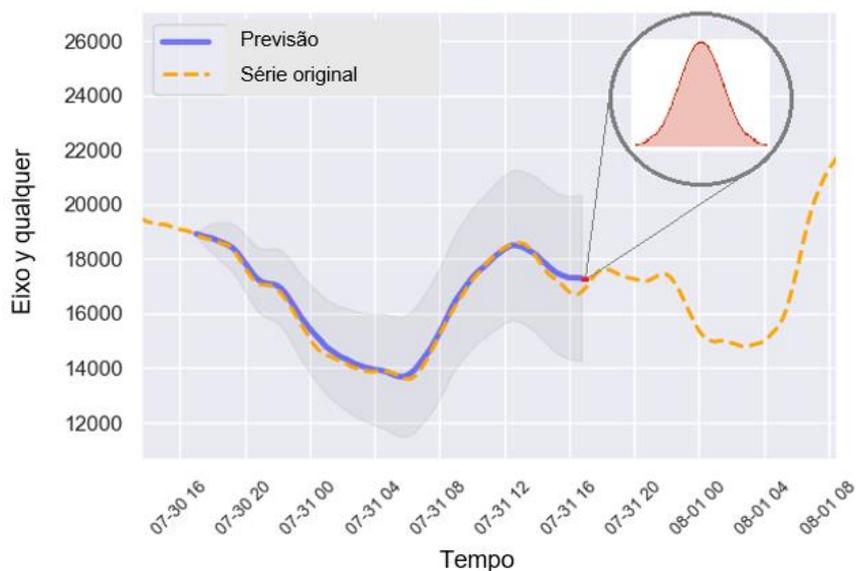
A **Erro! Autoreferência de indicador não válida.** mostra a série temporal gerada com base no processo AR (2) da Equação 11. Em cada regressão há um termo de erro que é caracterizado pela incerteza inerente do processo. O erro é representado por uma distribuição normal com média zero e variância constante

Figura 3: Série temporal descrita por um processo AR (2).



A Figura 4 mostra a distribuição de erros em torno de um valor previsto.

Figura 4: Distribuição de erro em torno da previsão, mostrando também o intervalo de confiança de 95% em torno da previsão



2.4.2 Modelos de média móvel

Semelhante ao modelo AR, que regride usando valores passados da série temporal, o modelo de média móvel (MA) usa erros como variáveis explicativas. O processo é dito ser um modelo MA de ordem q . MA (q) é descrito na Equação 12.

(12)

Coeficientes de q erros passados

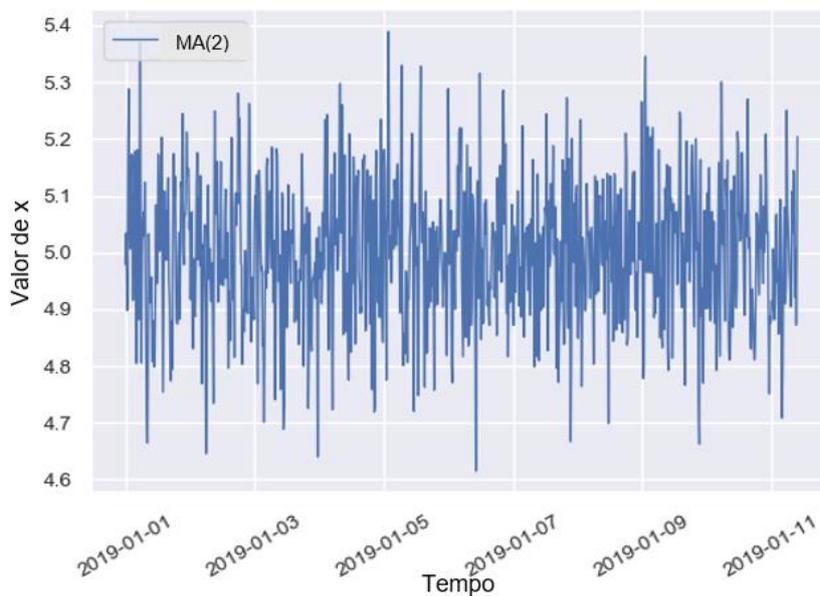
Erro defasado de q unidades

O modelo de média móvel de ordem 2, i.e., MA (2) é dado pela Equação 3

(13)

A Figura 5 apresenta um exemplo de um processo MA (2).

Figura 5: Série temporal descrita por um processo MA (2)



2.4.3 Modelos auto regressivos de médias móveis

O modelo ARMA é uma combinação do modelo AR e MA. Tal modelo usa valores passados da própria série temporal e valores passados de erros para descrever o valor atual do conjunto de dados. O processo é dito ser um ARMA (p, q) descrito pela Equação 14:

(14)

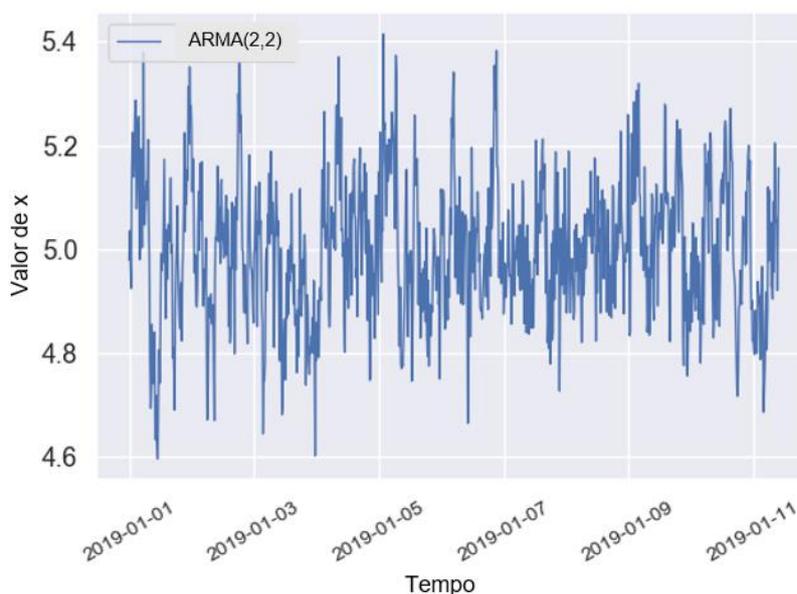
Os modelos ARMA podem ser representados de forma comprimida pela Equação 15

(15)

A Figura 6 apresenta o exemplo de um processo ARMA (2,2) baseado na equação 16 a seguir:

(16)

Figura 6: Série temporal descrita por um processo ARMA (2,2)



2.4.4 Modelo auto regressivo integrado de médias móveis

Os modelos ARMA só podem ser utilizados com séries temporais estacionárias; no entanto, na prática, a maioria das séries temporais não mostra o comportamento de estacionariedade. O modelo ARIMA consiste de uma união do modelo AR, do modelo MA e uma etapa de integração, usada para remover o componente de tendência (BROCKWELL, 1996). Os parâmetros dos modelos ARIMA são p , d e q ; portanto, escrito como $ARIMA(p, d, q)$ onde p , d , q são inteiros não negativos. O processo $ARIMA(p, d, q)$ é obtido pela integração temporal d vezes da série temporal $ARMA(p, q)$. Nos modelos ARIMA, uma série temporal não estacionária é tornada estacionária aplicando a diferenciação finita dos seus valores. O inteiro d controla o nível de diferenciação (VAGROPOULOS, CHOULIARAS, *et al.*, 2016). Para uma melhor compreensão do modelo, o operador ∇ e o operador de diferenciação ∇^d são introduzidos na Equação 17 e Equação 18:

(17)

Operador defasagem
 i Número da defasagem
 Observações passadas de acordo com i

(18)

Operador diferenciador de ordem 1

O modelo geral ARIMA (p, d, q) é descrito pela Equação 19 e em forma comprimida na Equação 20.

(19)

Derivada d da série temporal

(20)

Operador defasagem

Coeficientes AR

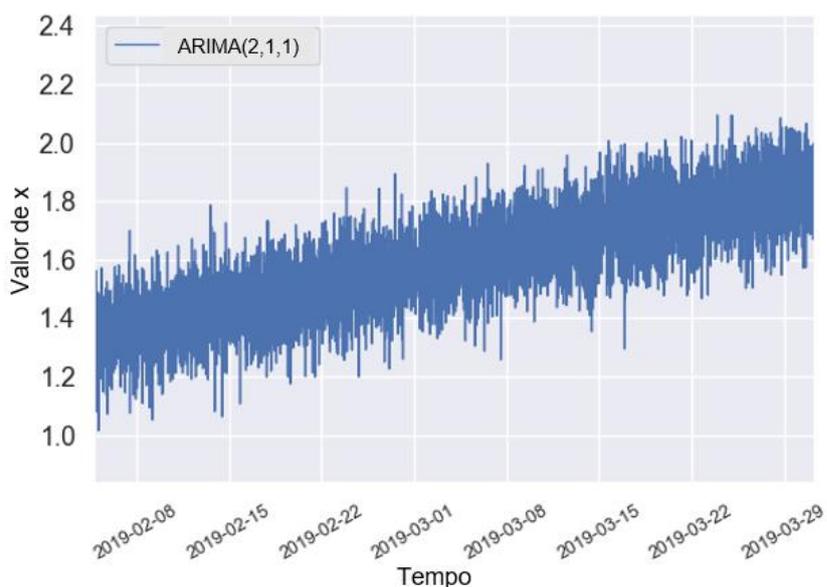
Coeficientes MA

Um exemplo de ARIMA (2,1,1) é expresso na Equação 21, que é apresentada na

Figura 7.

(21)

Figura 7: Série temporal descrita por um processo ARIMA (2,1,1)



2.4.5 Modelo auto regressivo integrado de média móvel sazonal

Os modelos SARIMA têm dois conjuntos de parâmetros, sendo eles, SARIMA $(p, d, q) \times (P, D, Q) S$, o pequeno d refere-se à diferenciação em relação aos valores precedentes e o grande D refere-se a diferenciação de ordem sazonal, usando S defasagens de forma a remover a sazonalidade. A diferença entre os modelos SARIMA e ARIMA é que os modelos ARIMA são usados para séries temporais com características não sazonais e o modelo SARIMA é destinado a séries temporais com características sazonais. Uma diferença sazonal de primeira ordem é a diferença entre uma observação e a correspondente observação do período sazonal anterior. Em um

modelo SARIMA, os termos AR e MA sazonais predizem usando valores e erros da série temporal com defasagens que são múltiplos de S (duração da sazonalidade). Por exemplo, com uma sazonalidade mensal, um modelo auto regressivo sazonal de primeira ordem usa para calcular e um modelo auto regressivo de segunda ordem usa e para calcular . Essa diferenciação é geralmente denotada novamente usando o operador de defasagem L . O modelo geral SARIMA $(p, d, q) \times (P, D, Q)_S$ é descrito pela Equação 22:

(22)

Coeficiente sazonal AR

Coeficiente não sazonal AR

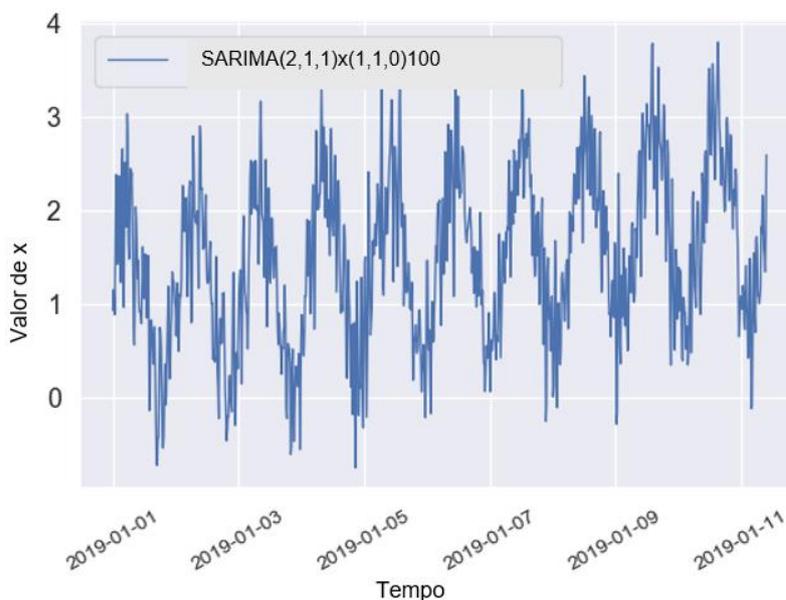
Coeficiente sazonal MA

Coeficiente não sazonal MA

Um exemplo de SARIMA $(2,1,1) \times (1,1,0)_{100}$ é expresso na Equação 23, que é apresentado na Figura 8:

(23)

Figura 8: Série temporal que representam um processo SARIMA $(2,1,1) \times (1,1,0)_{100}$



Os modelos SARIMAX funcionam de maneira muito semelhante aos modelos SARIMA. Nesse caso, os preditores exógenos são inseridos junto com a variável de previsão para realizar um melhor ajuste do modelo. A equação do modelo SARIMAX pode ser vista na Equação 24:

(24)

Coeficiente sazonal AR
 Coeficiente não sazonal AR
 Coeficiente sazonal MA
 Coeficiente não sazonal MA
 Coeficiente do termo exógeno
 Termo exógeno da série temporal

O modelo SARIMA pode ser transformado em qualquer outro modelo previamente documentado com o conjunto correto de parâmetros, ou seja, o SARIMA (p, d, q) $(0,0,0)$ é de fato o ARIMA (p, d, q) modelo e o ARIMA $(p, 0, q)$ é na verdade o modelo ARMA (p, q) (BROCKWELL, 1996).

2.5 Determinação da ordem dos modelos

2.5.1 Função de auto correlação e função de auto correlação parcial

Uma questão que surge é como determinar a ordem dos modelos estatísticos. Em outras palavras, até quantas defasagens de observações passadas é benéfico para aumentar a precisão do modelo? Por exemplo, a temperatura ambiente atual está fortemente ligada à temperatura de uma hora atrás, mas mostra uma correlação menor com a temperatura do mês passado. A função de auto correlação (FAC) e a função de auto correlação parcial (FACP) são ferramentas estatísticas que ajudam a encontrar o número apropriado de defasagens. A correlação de duas variáveis aleatórias é uma medida de quão próximas essas variáveis são de ter uma relação linear entre si.

A auto covariância com defasagem k é definida na Equação 25

(25)

Valor esperado da série temporal

Covariância da série temporal com respeito à defasagem k

O coeficiente de auto correlação com defasagem k é definido na Equação 26

(26)

Correlação de Pearson

Covariância da série temporal com respeito à defasagem k

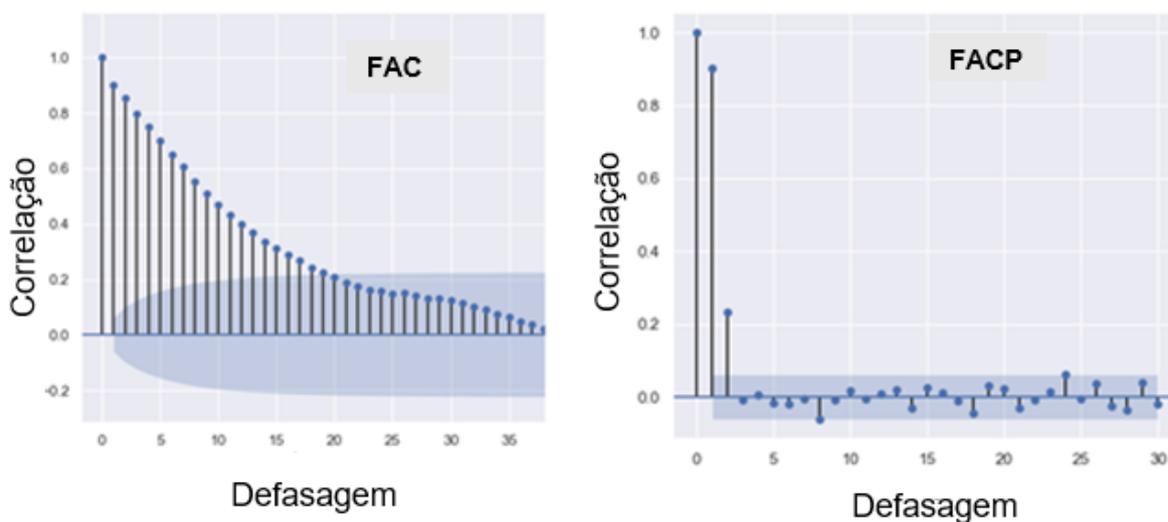
Valor médio da série temporal

Variância da série temporal

A partir da definição, é possível verificar que o coeficiente de auto correlação é adimensional e A FACP é a medida da correlação entre a observação atual e a observação defasada em k vezes, com a dependência linear entre as defasagens removidas (i.e. em defasagens $< k$) (BROCKWELL, 1996).

A FAC e a FACP são usadas para determinar se o sistema é governado por um modelo AR, um modelo MA ou uma combinação dos dois (ARMA). Essas funções também ajudam a identificar se a série temporal possui sazonalidade ou não. Figura 9 abaixo mostra um exemplo da FAC e da FACP (JOHN Y. CAMPBELL, 1997).

Figura 9: FAC e FACP de um processo de AR, onde o FAC diminui gradualmente e o FACP corta depois de p defasagens (BROCKWELL, 1996).



Se o sistema for modelado por um processo AR, o gráfico de auto correlação gradualmente diminui a zero e o gráfico de auto correlação parcial será cortado após p defasagens. Se o sistema for governado por um modelo MA, o gráfico de auto correlação será cortado após q defasagens e o gráfico de auto correlação parcial diminui gradativamente até zero. No modelo ARMA, o gráfico de auto correlação e o gráfico de auto correlação parcial diminuem gradualmente até zero. Esta informação está resumida na Tabela 2.

Tabela 2: Características de FAC e FACP de séries temporais estacionárias

Model	ACF	PACF
	Diminui gradualmente	Corta depois de p defasagens
	Corta depois de q defasagens	Diminui gradualmente
	Diminui gradualmente	Diminui gradualmente

Para o uso correto da ferramenta FAC e FACP, a série temporal deve ser estacionária. Isso significa que a auto correlação para qualquer defasagem específica é a mesma, independentemente de onde a série está no tempo. Não é difícil detectar a ordem de um modelo AR ou MA, no entanto com um sistema governado por ARMA o FAC e o FACP decaem para zero. Dessa forma, diante de um modelo ARMA, observar o gráfico FACP e o gráfico FAC não ajuda a encontrar valores de p e q .

2.5.2 Grid Search

Os modelos SARIMA possuem sete parâmetros, dos quais quatro são parâmetros sazonais. A maneira tradicional de definir os parâmetros através do uso de gráficos FAC e FACP era justificada quando o uso de computação era limitado e caro. Uma abordagem alternativa é utilizar o *Grid Search*, que é uma abordagem onde conjuntos de parâmetros possíveis são testados para descobrir qual modelo produz um erro de previsão menor (BOVAS ABRAHAM, 1983).

2.6 Medição do desempenho da previsão

2.6.1 Definição de erro de previsão

As etapas iniciais para a criação de um modelo de previsão são semelhantes entre todos os modelos. Inicialmente, o conjunto de dados deve ser dividido em conjunto de dados de treinamento e conjunto de dados de teste. O conjunto de treinamento é usado para a construção do modelo. Depois que o modelo é construído, é necessário avaliar a qualidade da previsão em relação ao conjunto de teste. Para obter um valor de precisão, é necessário realizar testes de desempenho. Esses testes são conhecidos como métricas de desempenho. Uma das métrica de desempenho utilizadas é a raiz média dos erros quadráticos, RMSE apresentado na Equação 27.

$$\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n} \quad (27)$$

Número de amostras

Observação no tempo t da série temporal

Observação no tempo t da previsão

Outro método para avaliar o desempenho da previsão é o erro percentual absoluto (EPA) apresentado na Equação 28.

$$\frac{\sum_{t=1}^n |y_t - \hat{y}_t|}{n} \quad (28)$$

Outro método para avaliar o erro de previsão é uma variação do erro percentual, mas ao invés de dividir o erro pela série temporal ele é dividido pelo máximo local, isso é feito para evitar a divisão por um número próximo de zero, nos casos em que a geração fotovoltaica é próxima de zero, tal formula é mostrada na Equação 29.

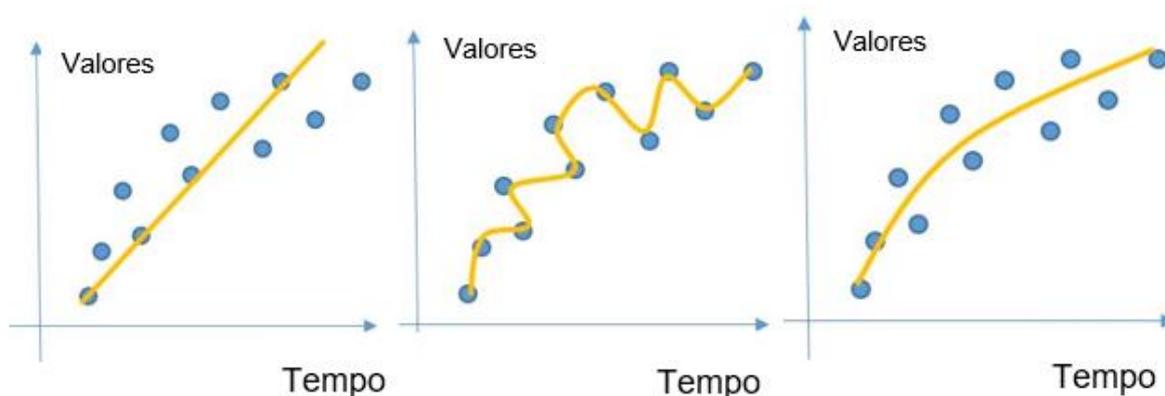
$$\frac{\sum_{t=1}^n |y_t - \hat{y}_t|}{\max(y_t)} \quad (29)$$

2.6.2 Validação cruzada

Diz-se que um modelo é *underfitted* quando o modelo não consegue capturar o comportamento fundamental do conjunto de dados de treinamento. Isso acontece quando o modelo não tem a complexidade necessária para descrever com precisão o relacionamento do conjunto de dados de treinamento, ou seja, quando se usa menos parâmetros que o necessário. Alternativamente, diz-se que um modelo é *overfitted* quando o modelo considera cada desvio, incluindo o ruído, nos dados de treinamento para ajustar seus coeficientes. O *overfitting* é muitas vezes o resultado de um modelo excessivamente complexo, quando o modelo tem um número excessivo de parâmetros, tendo um desempenho muito bom no conjunto de dados de treinamento,

e um desempenho ruim nos dados de teste. A Figura 10 apresenta o respectivo modelo *underfitted*, o modelo *overfitted* e o modelo otimizado.

Figura 10: Modelo *underfitted*, modelo *overfitted* e modelo otimizado respectivamente.



2.7 Conclusão

Quando se trata de realizar previsões usando métodos estatísticos, fica claro que na maioria dos casos a série temporal tem que ser manipulada para respeitar os requisitos estacionários, já que os processos naturais produzem séries temporais com tendências, componentes cíclicos e componentes sazonais. A alta capacidade de computação disponível hoje garante que os testes possam ser feitos com mais parâmetros e mais dados. Isso permite experiências com diferentes abordagens de manipulação, uma vez que os resultados dos modelos são adquiridos mais rapidamente do que era anteriormente. A partir deste capítulo, pode-se ver como a FAC e a FACP podem ser usadas para encontrar o período de sazonalidade. Desta forma é possível determinar quantas sazonalidades diferentes a série temporal é composta. Se houver mais de uma sazonalidade, será necessário pré-processar até restar no mínimo apenas uma sazonalidade, pois existe um modelo que pode representar uma série temporal sazonal. A FAC e a FACP podem ser usados para determinar os parâmetros do modelo AR e MA, embora não possam ser usados para encontrar os parâmetros de modelos mais complexos. A alternativa é usar o processo de *Grid Search*. O processo de *Grid Search* discrimina os modelos com baixos erros de previsão dos modelos com altos erros de previsão. Em outras palavras, a lista de

modelos, retornados do *Grid Search*, é uma coleção de modelos *overfitted*, e ótimos que, posteriormente, precisam ser avaliados para separar o modelo ótimo dos *overfitted*.

3 Conjuntos de dados de previsão e conjunto de dados auxiliares

3.1 Introdução

Com o aumento da capacidade de armazenamento e o aumento do desempenho dos computadores, é possível reunir mais dados com uma taxa de amostragem maior. Para garantir que o uso de dados auxiliares como preditores sejam úteis, tais dados devem apresentar algum grau de relação com os dados de previsão. Nesta seção, será explicado como os dados auxiliares foram avaliados, no sentido de descobrir quais deles poderiam descrever satisfatoriamente os dados de previsão.

Neste capítulo, encontra-se uma descrição das variáveis de previsão e das variáveis preditoras auxiliares. É informado a origem dos dados coletados, a taxa de amostragem e o intervalo temporal dos mesmos. Outra questão abordada neste capítulo trata da execução das técnicas de pré-processamento para garantir que as séries temporais respeitem os requisitos dos modelos. Para uma determinada característica na série temporal, há uma abordagem apropriada para isolar as características determinísticas da série temporal.

3.2 Variáveis de previsão

As ferramentas aqui desenvolvidas podem ser utilizadas para qualquer empresa transmissora da Alemanha, uma vez que é obrigatório a disponibilidade destes dados. A Amprion GmbH, um operador de sistema de transmissão (OST) baseado em Dortmund, é uma das quatro empresas de transmissão existentes na Alemanha. As empresas de transmissão, atuam na operação da rede de transmissão garantindo que a rede opere de forma segura, confiável sem abrir mão da eficiência energética. A área de controle da Amprion é visualizada na Figura 11. Quatro conjuntos de dados da Amprion foram usados para criar os modelos estatísticos de

previsão. Os conjuntos de dados utilizados são geração fotovoltaica, geração eólica, carga e carga vertical que variam de 01/01/2011 a 31/12/2016 com uma taxa de amostragem de 15 minutos. Em fevereiro de 2019, foi medido que a geração fotovoltaica teve uma capacidade instalada de 10343 MW e a geração eólica *onshore* teve uma capacidade instalada de 10330 MW. Tanto a geração fotovoltaica quanto a geração eólica podem ser encontradas no site da Amprion (AMPRION, 2019) e representam a quantidade total de potência fotovoltaica e eólica em MW da área de concessão da Amprion. A carga e a carga vertical estão também disponíveis ao público e podem ser obtidas na Plataforma Europeia de Operadores de Redes de Transporte de Eletricidade (ENTSO-E). (entsoe). A carga vertical é a soma, positiva ou negativa, de toda a energia transferida da rede de transmissão para a rede de distribuição através dos transformadores. A Tabela 3 apresenta um resumo de todas as variáveis usadas para a previsão.

Tabela 3: Variáveis usadas nas previsões.

Variáveis	Fonte	Resolução	Extensão
Geração solar	Site da Amprion	15-minutos	01/01/2011- 31/12/2016
Geração eólica	Site da Amprion	15-minutos	01/01/2011- 31/12/2016
Carga	Plataforma Entsoe	15-minutos	01/01/2011- 31/12/2016
Carga vertical	Plataforma Entsoe	15-minutos	01/01/2011- 31/12/2016

Figura 11: Área de controle da Amprion em roxo



Fonte: Amprion (2019)

3.2.1 Variáveis auxiliares

O *Deutscher Wetterdienst* (DWD) é uma empresa meteorológica alemã, que monitora o clima e as condições meteorológicas e fornece serviços meteorológicos para o público em geral. O DWD oferece acesso gratuito aos dados climáticos através do Centro de Dados Climáticos (CDC). O CDC permite acesso via servidor FTP, onde os dados são atualizados automaticamente ou por download direto do servidor CDC. Globalmente, o DWD opera mais de 3000 estações meteorológicas, onde em território alemão existem 179 estações. Uma lista de todas as condições meteorológicas, com

localização, altitude e outras características, pode ser encontrada no site da DWD. (Deutscher Wetterdienst).

Usando as coordenadas geográficas dos transformadores da Amprion e as coordenadas geográficas das estações meteorológica DWD (Deutscher Wetterdienst), foi possível definir um critério de proximidade para definir quais estações meteorológicas estão localizadas dentro da área de controle da Amprion. Deste total, 39 das 179 estações meteorológicas na Alemanha foram selecionadas, por estarem na área de concessão da empresa de transmissão. De cada estação meteorológica, os seguintes dados históricos com uma resolução de 1 hora foram adquiridos. A Tabela 4 descreve as informações gerais sobre as variáveis auxiliares.

Tabela 4: Estrutura das variáveis preditoras.

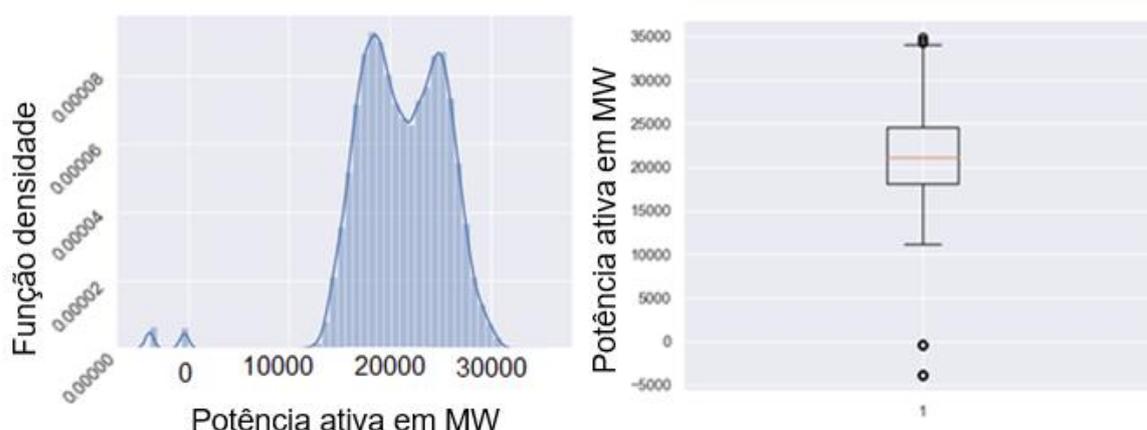
Variável	Fonte	Resolução	Extensão
Irradiação solar	DWD	60-minutos	01/01/2011- 31/12/2016
Temperatura	DWD	60-minutos	01/01/2011- 31/12/2016
Umidade	DWD	60-minutos	01/01/2011- 31/12/2016
Precipitação	DWD	60-minutos	01/01/2011- 31/12/2016
Velocidade do vento	DWD	60-minutos	01/01/2011- 31/12/2016
Direção do vento	DWD	60-minutos	01/01/2011- 31/12/2016

Um procedimento de interpolação foi aplicado nos dados meteorológicos, a fim de manter a mesma resolução de 15 minutos, conforme visto nas variáveis elétricas adquiridos no site da Amprion.

3.3 Preparação de variáveis para respeitar os critérios dos modelos

Antes de fornecer a série temporal para os modelos, é necessário realizar a manipulação da série temporal para garantir que respeitem os requisitos corretos propostos na abordagem estatística. A primeira etapa de preparação de dados é a remoção de *outliers*. Um *outlier* é uma observação que se desvia significativamente da média da amostra. Os *outliers* podem ser oriundos de um erro de medição ou podem ser uma observação genuína de valores extremos. Existem vários métodos para facilitar a descoberta de *outliers*; as mais simples contam com métodos de visualização. Dois gráficos diferentes foram usados para encontrar os *outliers*, o gráfico de distribuição e o *boxplot* como mostrado na Figura 12.

Figura 12: Gráfico de distribuição e boxplot do conjunto de dados de carga.



O *boxplot* é usado para verificar graficamente quais das observações estão fora do intervalo do quartil inferior e superior. O gráfico de distribuição é usado para se ter uma ideia de quão numerosos os *outliers* são estatisticamente - um *outlier* que é estatisticamente numeroso pode ser um valor real. Depois de encontrar os *outliers*, outra decisão importante a fazer é como substituir os *outliers* por valores razoáveis. Se os *outliers* forem isolados entre duas observações válidas, uma simples interpolação poderá ser feita. Em casos em que muitos *outliers* são agrupados, as informações de tendência e sazonalidade são usadas para substituir os *outliers*.

Além disso, um conjunto de dados alternativo composto pela suavização do sinal do conjunto de dados original é criado. A ideia por trás desse procedimento é remover a influência da amostragem, já que em situações reais a magnitude das observações ocorre de forma constante e geralmente suave. Estas duas séries temporais são então levados em conta para fazer as previsões, os dados originais e os dados suavizados através de um filtro. O objetivo da suavização é remover os valores que mudam rapidamente do sinal considerando que o sinal não está distorcido. O filtro executa um ajuste polinomial com uma janela de quatro unidades, duas unidades de tempo antes e duas unidades de tempo após o valor desejado a ser calculado.

3.4 Conjunto de dados de geração fotovoltaica

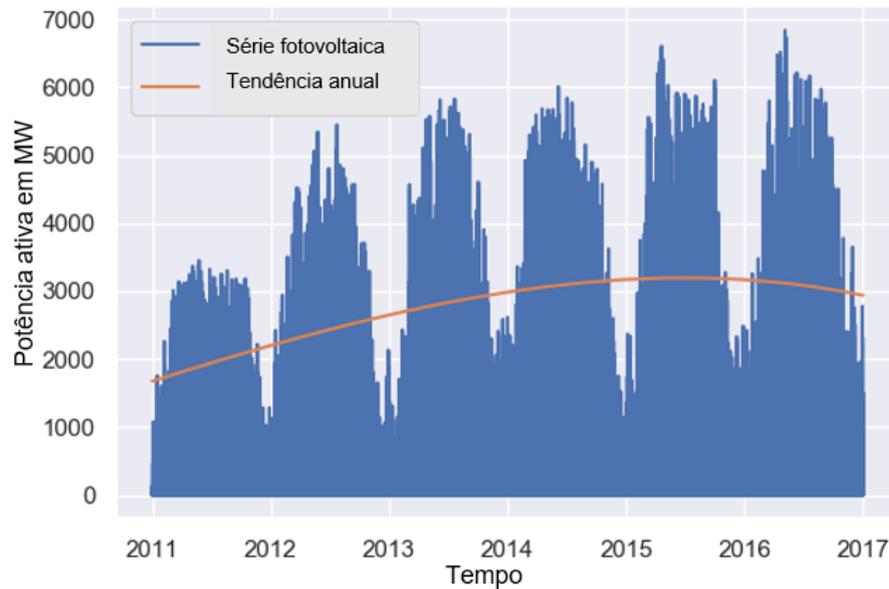
A série temporal original de geração fotovoltaica apresentada na Figura 13 revela uma tendência crescente ao longo dos anos. Esse padrão crescente é explicado pela capacidade instalada cada vez maior de geradores fotovoltaicos. Outro padrão que é muito evidente no conjunto é a sazonalidade da energia gerada. A tendência e a sazonalidade mostram que esta série temporal não respeita as diretrizes de estacionariedade. O procedimento de remoção destas características é ajustar as funções determinísticas à tendência anual que modela a crescente instalação de painéis fotovoltaicos e a sazonalidade diária e/ou anual. O padrão de sazonalidade modela o movimento de revolução da terra ao redor do sol. A modelagem do componente de tendência da geração fotovoltaica é realizada com um ajuste polinomial usando os valores diários máximos de geração fotovoltaica. A Equação 30 modela a tendência determinística da série temporal de geração fotovoltaica.

(30)

Componente da tendência

Na Figura 13 o conjunto de dados fotovoltaicos original com a função determinística da tendência anual é ilustrado.

Figura 13: Séries temporais fotovoltaicas com tendência



Seguindo o processo de obtenção da série temporal mais próxima à estacionariedade, a série temporal original é então dividida pela tendência anual determinística. A série temporal resultante não depende mais da quantidade de painéis fotovoltaicos instalados, sendo uma função apenas das condições climáticas, bem como dos movimentos de rotação e revolução. Para modelar o componente determinístico sazonal, um ajuste senoidal é realizado com a série temporal resultante. A Equação 31 é a equação geral do ajuste senoidal e a Equação 32 são os coeficientes calculados para este caso particular.

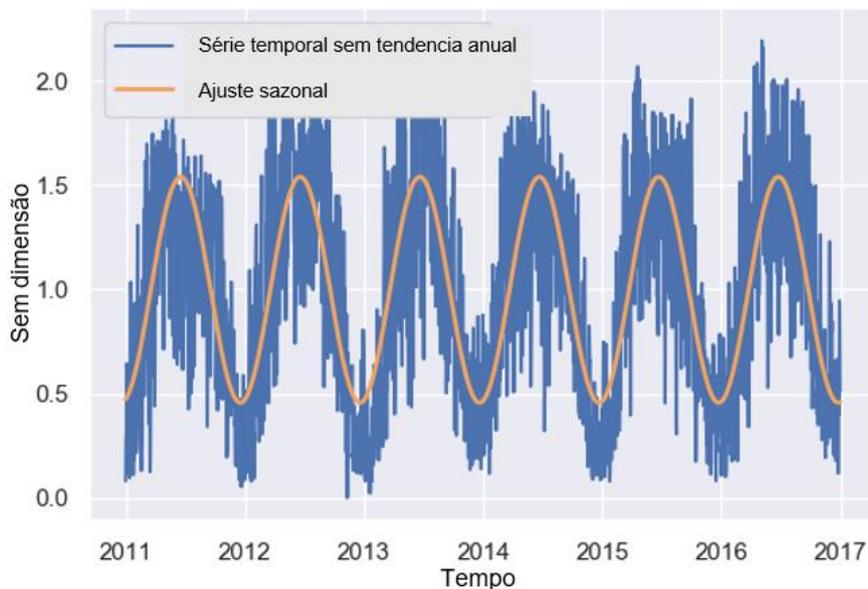
(31)

Componente da tendência
 Amplitude da função senoidal
 Frequência da função senoidal
 Defasagem em ângulo
 Valor médio

(32)

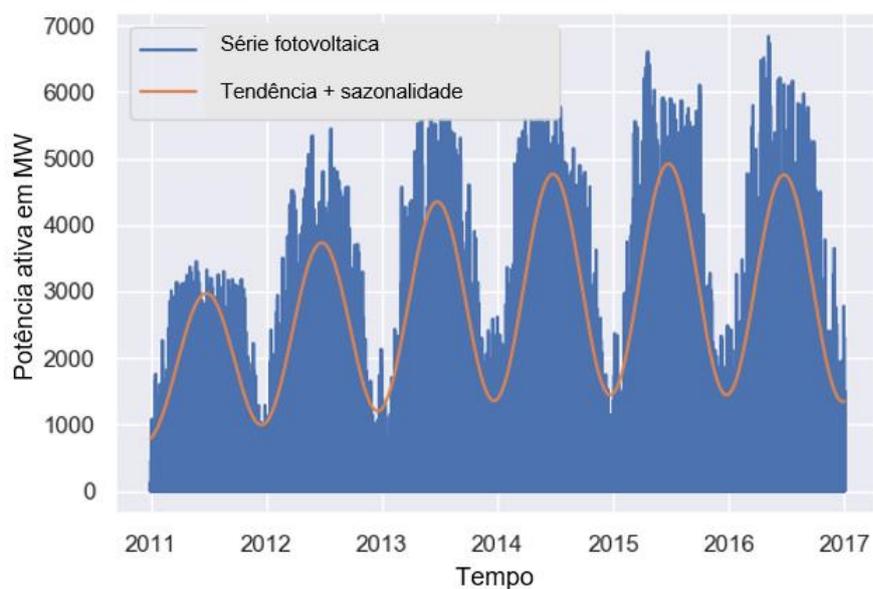
Figura 14 representa a dependência do conjunto de dados das estações do ano.

Figura 14: Sazonalidade de séries temporais fotovoltaicas



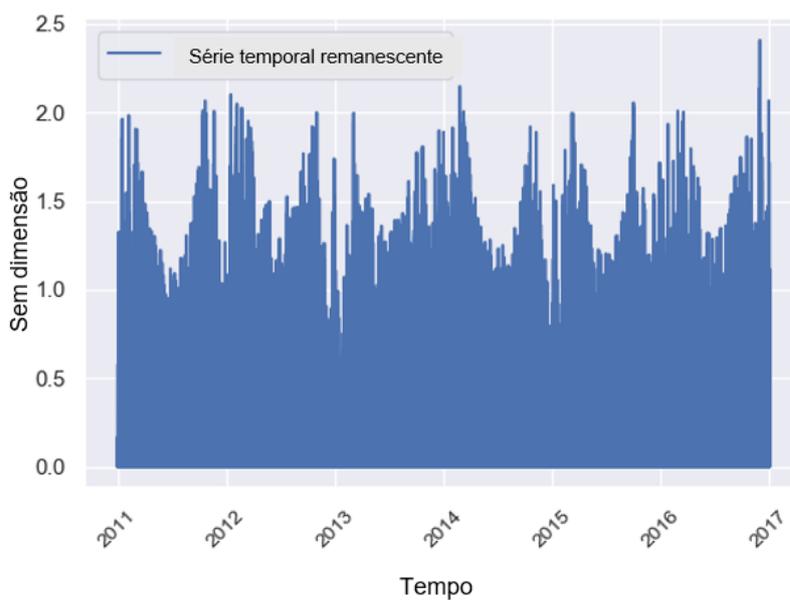
O conjunto de dados restante é então modelado usando uma função determinística e sazonalmente dependente. A Figura 15 apresenta a multiplicação da função de tendência anual com a função sazonal e como ela se relaciona com o conjunto de dados original em uma forma visual.

Figura 15: Multiplicação de tendência com sazonalidade sobreposta à série temporal original



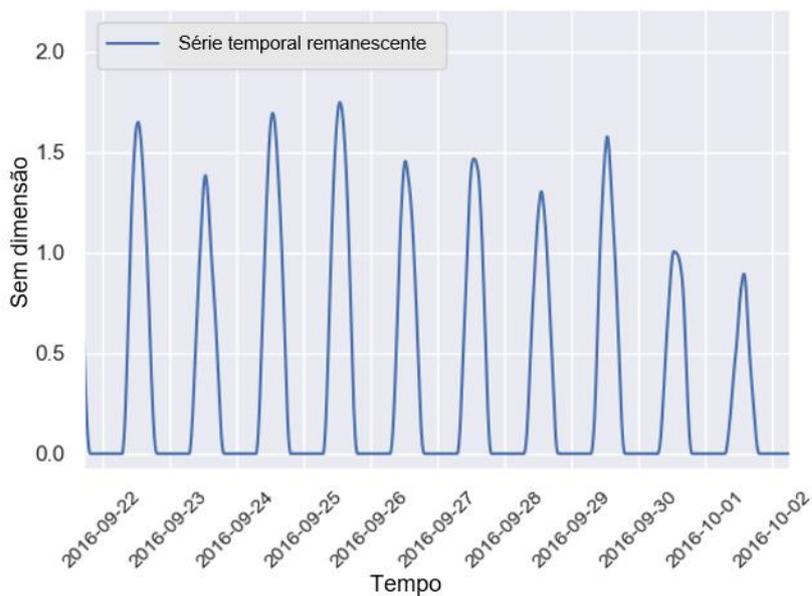
A série temporal fotovoltaica original é então dividida pela função que modela a tendência e a sazonalidade juntas. O resultado do processo é mostrado na Figura 16.

Figura 16: Séries temporais fotovoltaicas sem tendência e sem sazonalidade anual



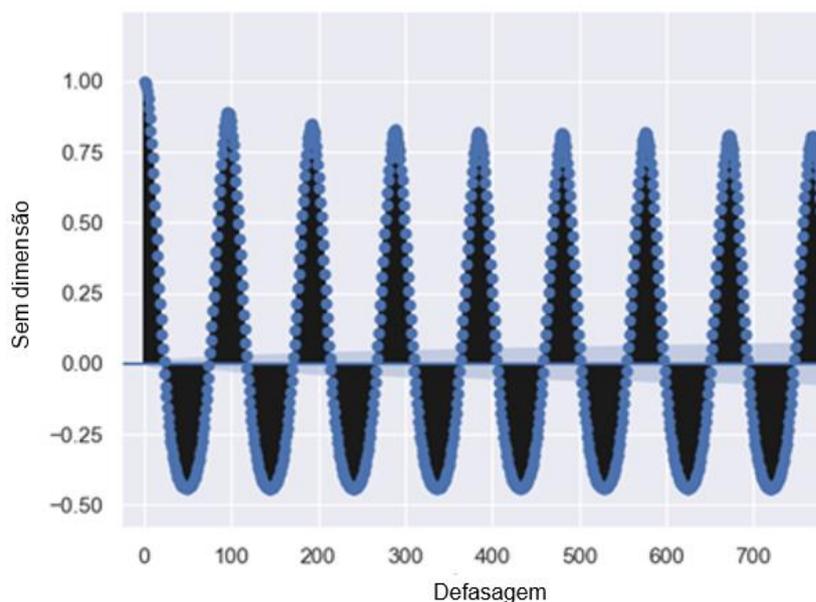
Através de uma inspeção visual, pode-se ver que as séries temporais ao longo dos anos não mostram uma tendência crescente e uma sazonalidade anual bem pronunciada. Ao ampliar a imagem, a sazonalidade diária pode ser observada, como mostra a Figura 17.

Figura 17: Série temporal fotovoltaica de sazonalidade diária



Não é necessário realizar uma remoção de sazonalidade diária, uma vez que existe um modelo que pode atender esta situação, o modelo SARIMA (X). Ao traçar a função de auto correlação, conforme ilustrado na Figura 18, pode-se ver os picos da função de auto correlação. Isso mostra que ainda há um forte componente sazonal na série temporal; - com o seu pico em múltiplos de 96, o número de amostras por dia, quatro amostras por hora em um período de 24 horas.

Figura 18: Gráfico da função de auto correlação da série temporal resultante



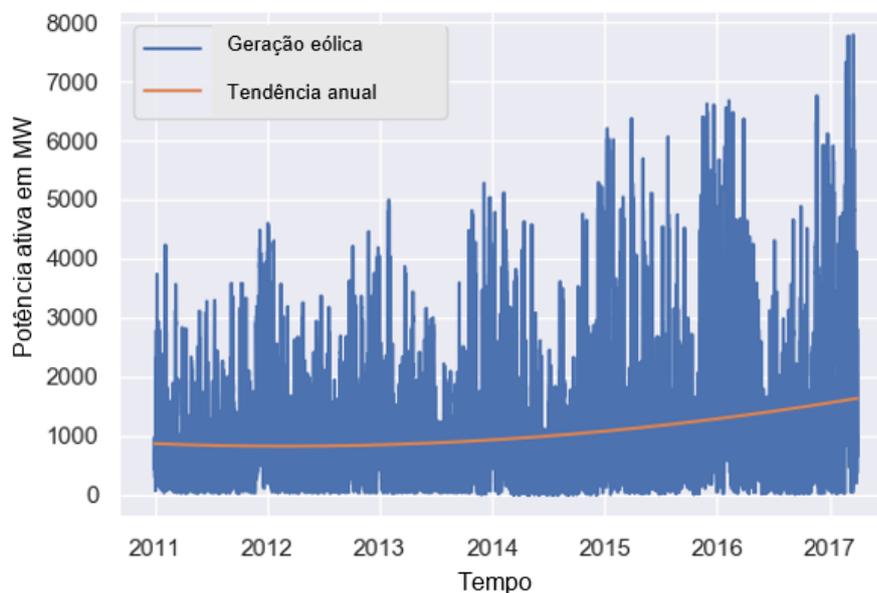
3.5 Conjunto de dados de geração eólica

Semelhante à série temporal de geração de energia fotovoltaica, a série temporal de geração eólica também mostra uma forte tendência anual, descrito pelo aumento das turbinas eólicas instaladas ao longo dos anos. Para encontrar a função determinística que modela o aumento de turbinas eólicas, um ajuste polinomial é realizado usando a média da série temporal com uma janela de um ano. A tendência anual é então modelada por uma equação de segunda ordem apresentada pela Equação 33.

(33)

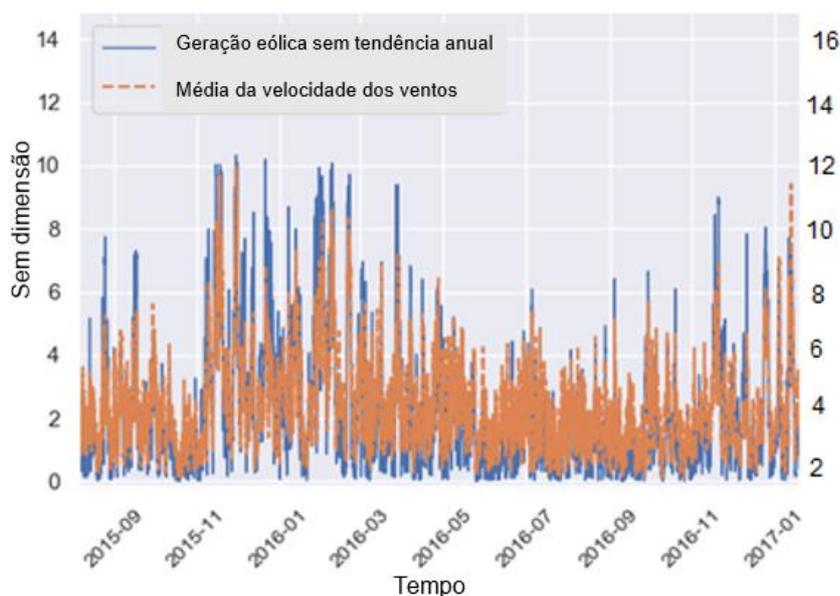
Figura 19 mostra o conjunto de dados de geração eólica com a tendência determinística anual.

Figura 19: Séries temporais de geração de energia eólica com sua respectiva tendência



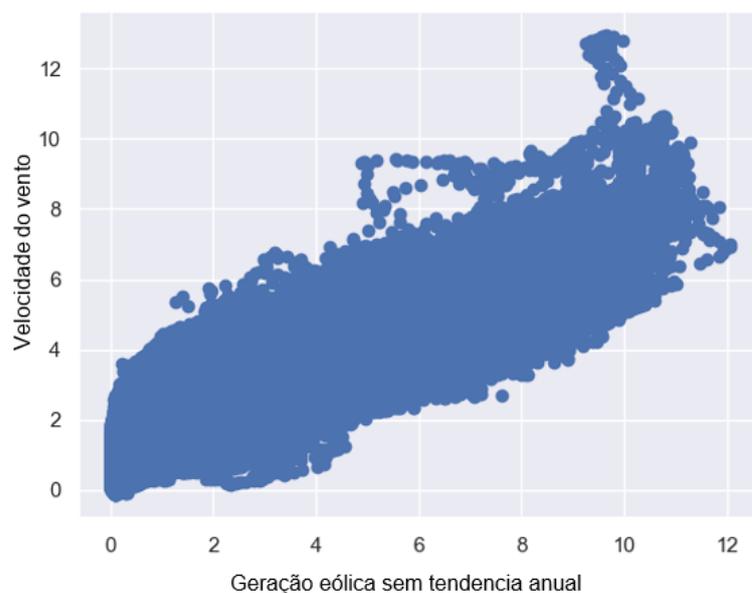
A série temporal de geração eólica original foi então dividida pela tendência anual, a fim de remover a influência da crescente oferta de turbinas eólicas. Na Figura 20, o conjunto de dados residual resultante e a velocidade média do vento de todas as estações meteorológicas nas áreas de controle da Amprion são mostradas.

Figura 20: Similaridades entre os dados de velocidade média do vento exógeno e a série temporal de geração eólica resultante



A julgar pelo gráfico de dispersão mostrado na Figura 21, é evidente que os dados exógenos de velocidade de vento são bons estimadores para a geração de energia eólica.

Figura 21: Gráfico de dispersão mostrando o grau de correlação entre a média da velocidade de ventos na região da Amprion e a energia eólica gerada na mesma região.



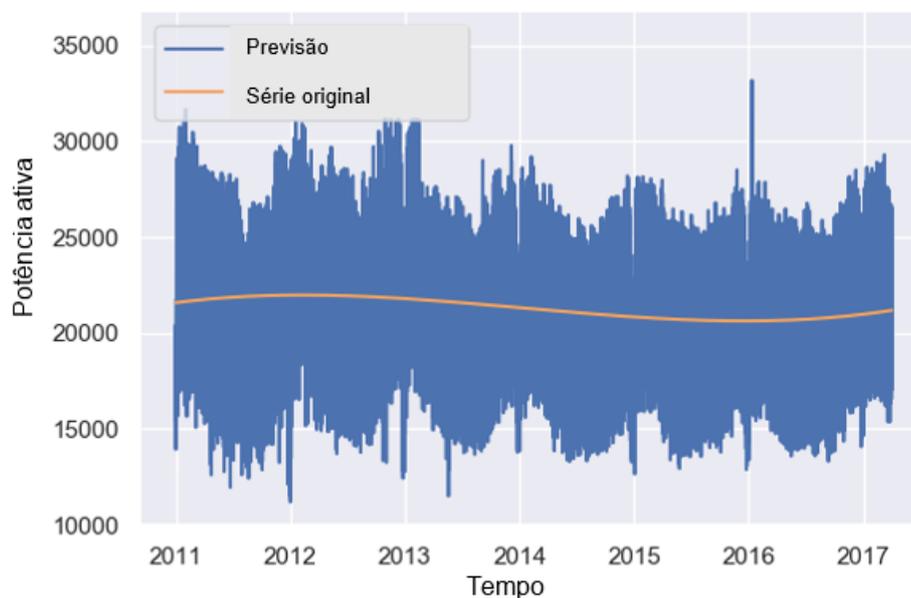
3.6 Série temporal da carga

A tendência determinística da série temporal de carga é obtida pela média móvel tomando como janela o período de um ano. A resultante deste processo é obtida através de um ajuste polinomial de ordem 3. A equação é apresentada pela Equação 34:

(34)

Figura 22 mostra o gráfico de carga junto com sua tendência determinística calculada.

Figura 22: Tendência anual do conjunto de dados de carga sobreposta à série temporal da carga

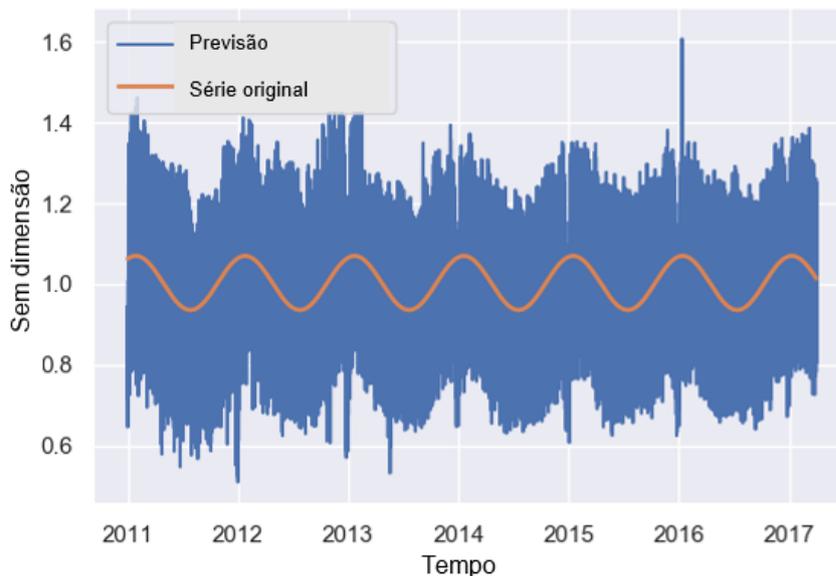


A série temporal de carga possui uma sazonalidade anual, oriunda de um consumo de energia maior no inverno e um consumo menor de energia no verão. A série temporal também mostra uma sazonalidade semanal criada pelos hábitos semanais semelhantes da população em geral. Depois de remover a tendência anual do conjunto de dados, esses dois tipos de sazonalidade periódica se tornaram muito evidentes. Uma função determinística senoidal é então ajustada modelando os efeitos do verão e do inverno sobre as séries temporais como mostrado na Equação 35:

(35)

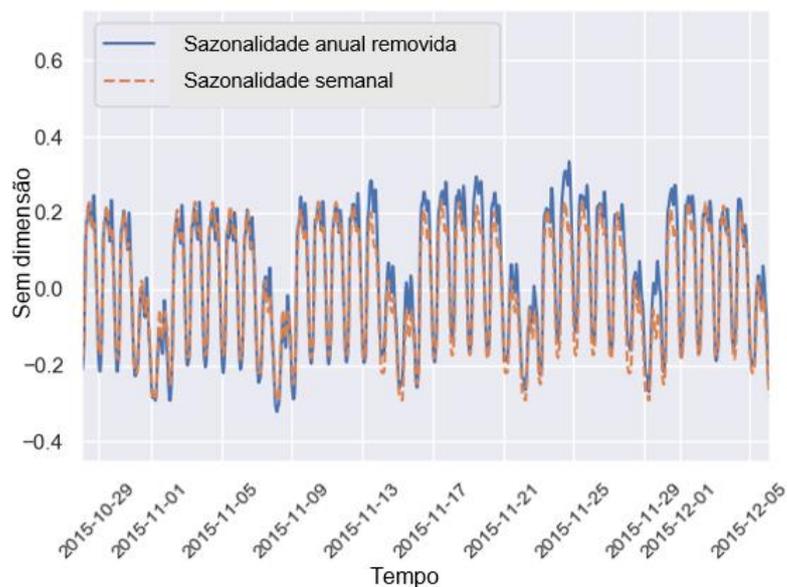
Figura 23 apresenta o conjunto de dados remanescente após remover a tendência e o ajuste senoidal que modela a sazonalidade anual de inverno e verão.

Figura 23:A sazonalidade anual do conjunto de dados de carga se sobrepõe à série temporal da carga



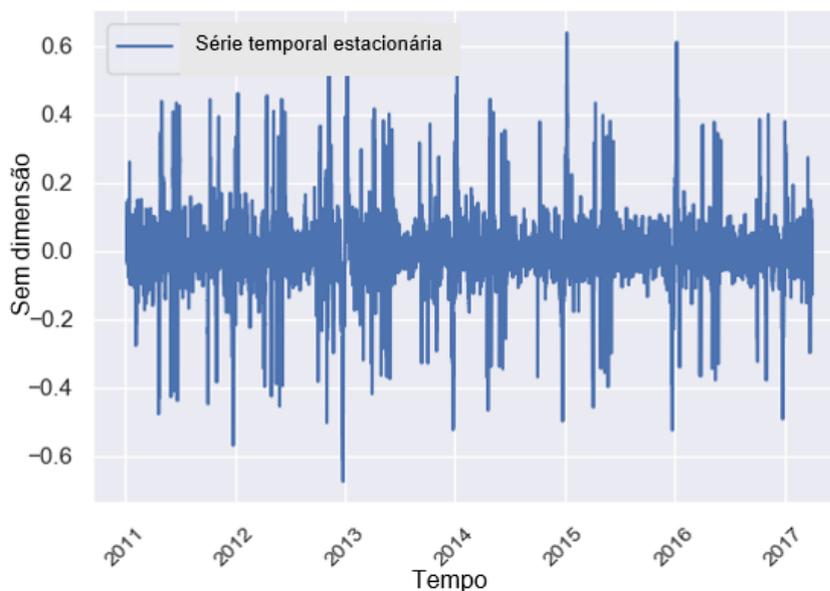
A série temporal restante possui uma forte sazonalidade semanal. Essa sazonalidade está relacionada aos hábitos das pessoas, por exemplo, o hábito semanal da maioria das pessoas de acordar por volta das 7h da manhã, almoçar por volta das 12h e ir para casa por volta das 18h nos dias de trabalho, já no fim de semana acorda-se mais tarde e almoça-se mais tarde. Para remover a influência da sazonalidade semanal, um processo de diferenciação é realizado usando a duração da sazonalidade semanal (ou seja, 672 defasagens), que corresponde ao período de uma semana. A Figura 24 mostra a sazonalidade semanal junto com a carga média semanal. A carga semanal média é a média dos valores de todos os padrões de carga semanais observados em todo o conjunto de dados de carga.

Figura 24: Comparação do conjunto de dados de carga quando sua sazonalidade anual foi removida com o padrão de sazonalidade da carga semanal.



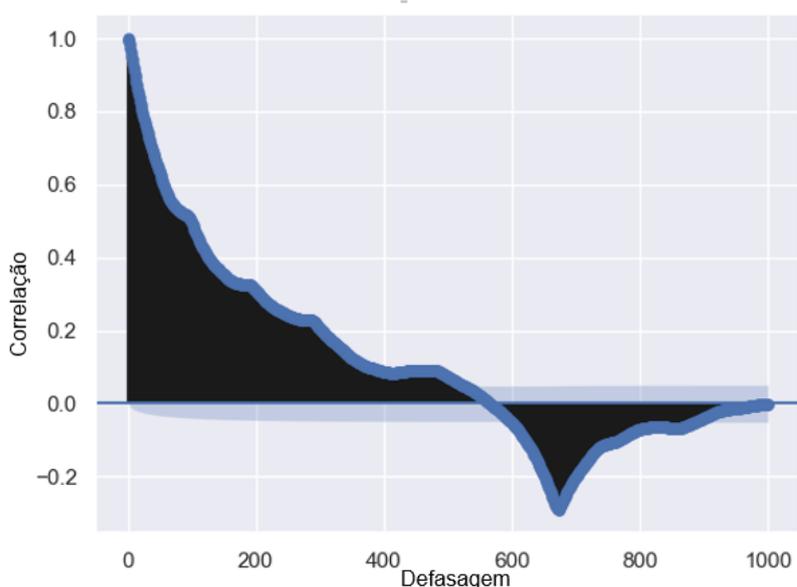
A subtração de séries temporais com as mesmas séries temporais defasadas em uma semana produz as séries temporais resultantes que não mostram uma sazonalidade semanal. O conjunto de dados residual é apresentado na Figura 25.

Figura 25: Série temporal residual da carga



Com o uso da FAC observa-se que a série temporal ainda possui uma dependência anual causada pela carga reduzida entre o Natal e a véspera de Ano Novo, bem como uma forte relação com seus valores passados, conforme indicado na Figura 26.

Figura 26: FAC da série temporal resultante da carga



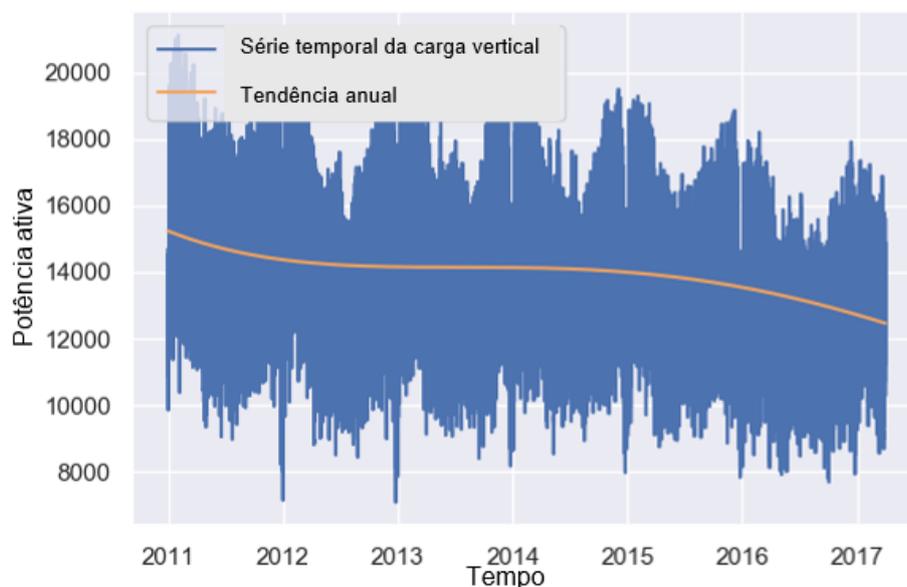
3.7 Série temporal de carga vertical

A carga vertical mostra muitas semelhanças com a série temporal da carga e o procedimento de pré-processamento é feito de maneira análoga. Inicialmente, a média móvel com janela de um ano é realizada. A resultante deste procedimento é a evolução temporal da tendência. A tendência é então modelada usando um ajuste polinomial de ordem 3, como mostrado na Equação 36.

(36)

Figura 27 mostra o componente de tendência da série temporal juntamente com a série temporal da carga vertical.

Figura 27: Séries temporais de carga vertical com a linha de tendência anual

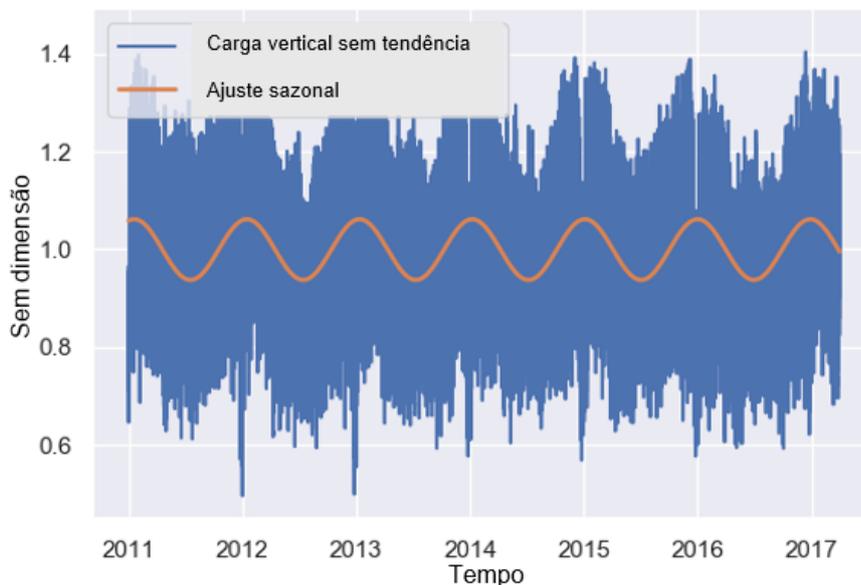


Para remover a influência da tendência anual, a série temporal original é dividida pela tendência determinística. A série temporal resultante é influenciada pela sazonalidade anual e sazonalidade semanal. Para modelar a sazonalidade anual da série temporal remanescente foi realizado um ajuste senoidal, a equação é mostrada na Equação 37.

(37)

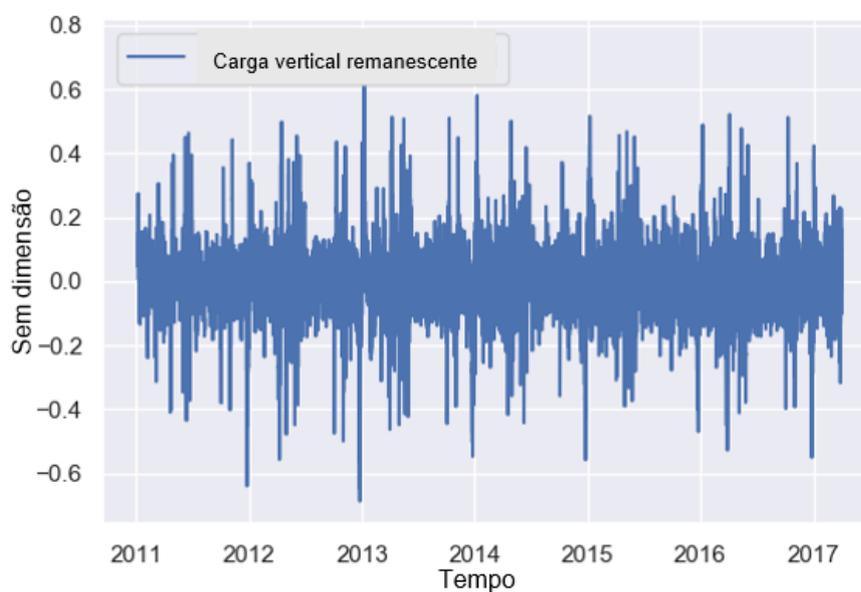
Figura 28 representa as séries temporais restantes com o ajuste senoidal modelando a sazonalidade anual.

Figura 28: A sazonalidade anual da série temporal de carga vertical sobreposta à série temporal da carga vertical.



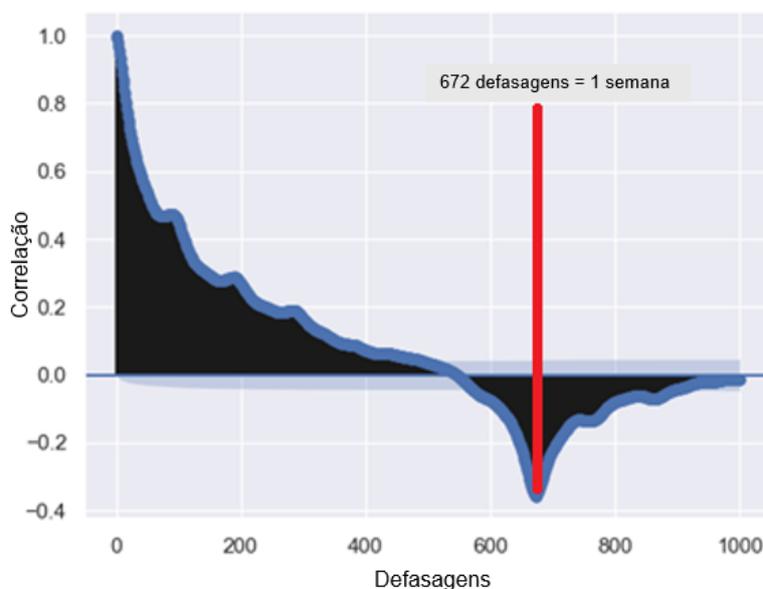
A série temporal restante tem uma forte sazonalidade semanal. Para remover essa sazonalidade, é realizado um processo de diferenciação. Para remover a sazonalidade semanal, a série temporal remanescente é subtraída por si própria, com 672 passos de defasagem, o que corresponde a uma semana. A Figura 29 mostra a série temporal residual.

Figura 29: Séries temporais de carga restante



Ao olhar o gráfico FAC na Figura 30, fica claro que o dia atual tem uma dependência nos dias anteriores e também na semana anterior, uma vez que o pico negativo ocorre no intervalo 672.

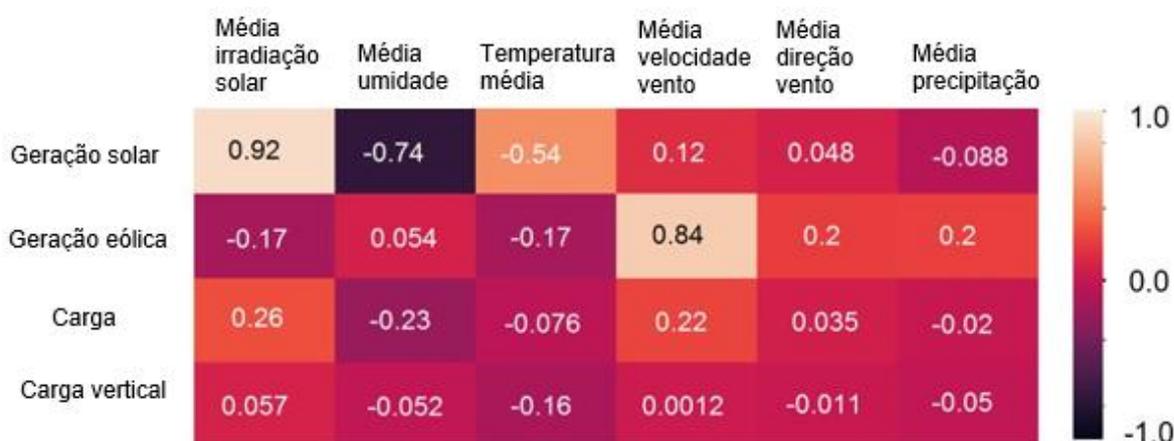
Figura 30: FAC dos resíduos da série temporal da carga vertical



3.8 Seleção de dados exógenos como preditores

Um preditor exógeno auxiliar é útil quando tem relação com a variável de previsão. Para inferir quais preditores têm uma relação mais forte com as variáveis de previsão, uma análise da correlação das variáveis de previsão, com a tendência removida, com relação às variáveis preditoras auxiliares é realizada. As variáveis de previsão sem tendência anual são usadas para essa análise, uma vez que o aumento da capacidade de geração fotovoltaica e eólica não se relaciona a nenhuma condição climática descrita no conjunto de dados dos preditores. A medida de correlação é um número entre -1 e 1. Quanto maior a magnitude absoluta da correlação, maior é o vínculo entre o preditor e as variáveis de previsão. O resultado da análise é mostrado na Figura 31.

Figura 31: Mapa de calor mostrando a correlação dos valores previstos em relação à média de cada preditor



A partir da análise da correlação é possível verificar que a geração fotovoltaica está fortemente correlacionada com a irradiação solar média, e com um grau de correlação menor, com a umidade média. A partir da análise também é possível verificar que a geração eólica está fortemente relacionada à velocidade média do vento, mas não fortemente relacionada com a direção do vento.

3.9 Conclusão

Tanto o gráfico da FAC quanto o da FACP são úteis para encontrar a sazonalidade no conjunto de dados e quantos períodos diferentes de sazonalidade a série temporal possui. A manipulação de dados de pré-processamento foi capaz de descrever os componentes de tendência e os componentes de sazonalidade com funções determinísticas. Essas funções determinísticas serão usadas para reconstruir os conjuntos de dados de previsão após uma previsão ser feita usando a parte residual da série temporal. De cada estágio do pré-processamento, o gráfico FACP e FAC poderia ser realizado para dar uma dica caso os componentes sazonais ou de tendência foram removidos. O conjunto de dados fotovoltaicos, o conjunto de dados de carga e o conjunto de dados de carregamento vertical tiveram seu componente sazonal modelado. O componente sazonal de vento não foi modelado, pois a sazonalidade anual da geração eólica não é muito pronunciada. A análise de correlação dos preditores em relação às variáveis de previsão indica que existe a

maior correlação entre a geração fotovoltaica e a irradiação solar média, bem como com a umidade média. Outra correlação considerável existe entre a geração eólica e a velocidade média do vento. Como resultado da Figura 31, a irradiação solar será utilizada como um preditor auxiliar para a geração fotovoltaica. A velocidade do vento será usada como um preditor auxiliar para a geração eólica. Tanto a carga quanto a carga vertical não mostram uma correlação muito intensa com quaisquer preditores. Portanto, para geração fotovoltaica e geração eólica, a família SARIMAX de modelos será usada, enquanto que para carga e carga vertical, o modelo da família ARMA será usado.

4 Desenvolvimento de modelos

4.1 Introdução

Este capítulo discute sobre a abordagem aplicada para encontrar o melhor método de previsão. Após os dados de previsão serem pré-processados, a parte residual deve ser aplicada ao modelo apropriado. Uma vez que a partir do gráfico FAC e FACP não é possível atribuir o modelo apropriado, uma coleção de modelos prováveis é designada para teste, isto é descrito como o processo de *Grid Search*. Depois de coletar os melhores modelos de previsão, ou seja, os que apresentam menores erros, é necessário separar o modelo ótimo dos modelos *overfitted*. Esse estágio, é chamado de validação cruzada. Este capítulo descreve detalhadamente o procedimento de *Grid Search* e validação cruzada, os resultados destes procedimentos e como eles foram implementados.

4.2 Implementação de métodos estatísticos para previsão de séries temporais

Os métodos estatísticos de previsão foram desenvolvidos usando a linguagem de programação *Python*. A linguagem de programação *Python* é uma linguagem de programação de propósito geral (não específico) de alto nível que vem crescendo em popularidade nos últimos anos devido ao seu uso em computação científica, *data mining*, e *machine learning*. O *Spyder* é uma plataforma IDE (ambiente de desenvolvimento integrado) de código aberto onde os códigos são escritos.

A linguagem de programação *Python* permite que o usuário usufrua de diversos pacotes e bibliotecas. Esses pacotes são códigos que possuem uma funcionalidade específica e podem ser instalados gratuitamente. Algumas das bibliotecas mais importantes utilizadas com sua respectiva funcionalidade estão listadas na Tabela 5.

Tabela 5: Bibliotecas relevantes usadas

Bibliotecas	Função
Pandas	Manipulação de dados e análise de dados
Numpy	Provê suporte a matrizes e vetores multidimensionais grandes
Statsmodel	Modelos estatísticos
Datetime	Manipulação de datas e tempo
Sklearn	Algoritmos de classificação e regressão
Scipy	Computação científica e técnica

Seguindo a abordagem de *Grid Search*, os parâmetros possíveis são primeiramente definidos para serem usados com o método SARIMAX. O modelo *Python* engloba os modelos AR, MA, ARMA, ARIMA e SARIMA - com ou sem dados auxiliares exógenos.

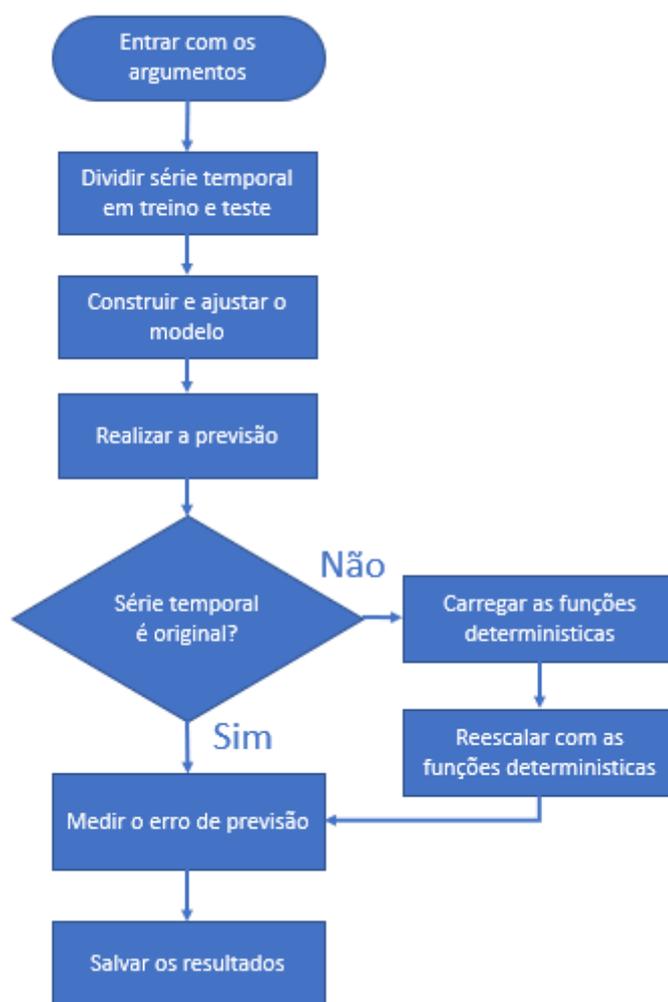
A estrutura do código pode ser dividida em oito seções conforme mostrado na Figura 32:

- Escolha dos conjuntos de dados
- Divisão dos conjuntos de dados
- Construção do modelo
- Ajuste do modelo
- Previsão
- Reescalamento

Medição de erro

Salvamento de resultados

Figura 32: Fluxograma de código SARIMAX



Em geral, quatro conjuntos de dados diferentes foram usados por variável de estudo. As séries temporais utilizadas com *Grid Search* foram:

Série temporal original

Série temporal suavizada

Série temporal estacionária feito com valor máximo diário (somente no caso fotovoltaico)

Série temporal estacionária feito com desvio padrão (somente no caso fotovoltaico)

A suavização de sinal executa um ajuste polinomial com uma janela de quatro unidades no tempo. Esta ferramenta foi usada considerando que a amostragem discreta causa a não suavidade na série temporal. No caso da série temporal fotovoltaica ficou claro como a série temporal suavizada representava uma situação que se assemelha mais à realidade do que a versão original da série temporal. A série temporal estacionária feito com o valor máximo diário foi obtido modelando uma função determinística sazonal usando o valor máximo diário de geração fotovoltaica. A outra série temporal para geração fotovoltaica foi feita modelando a função determinística sazonal usando o desvio padrão série temporal fotovoltaica.

Como o processo de ajuste do modelo (onde os coeficientes são calculados) é muito demorado, três diferentes períodos de tempo foram adotados para serem usados no treinamento do modelo. A intenção desta divisão é examinar como o aumento do conjunto de dados de treinamento contribui para um menor erro de previsão.

Depois que o modelo é criado, é realizado uma previsão de 10 passos para o futuro, equivalendo a 2,5 horas. A previsão, se necessário, é reescalada usando as funções determinísticas criadas no processo para torná-las estacionárias. Os dados resultantes são então comparados com o conjunto de dados original, o erro é medido e salvo junto com a previsão em um arquivo MATLAB.

Os modelos SARIMAX possuem sete parâmetros diferentes, que podem assumir valores diferentes: os parâmetros de ordem (p , d , q), o parâmetro de ordem sazonal (P , D , Q) e a sazonalidade (S). O método determina qual algoritmo iterativo de resolução é usado para encontrar as raízes da função de otimização. O algoritmo de resolução "nm" significa *Nelder-Mead*, "bfgs" significa *Broyden-Fletcher-Goldfarb-Shanno* (BFGS) e "Powell" significa o método de *Powell* modificado. O argumento de tendência no modelo determina se o modelo gerado não mostra nenhuma tendência (n) ou uma tendência constante (c), uma tendência linear (t) ou uma tendência

constante e linear (ct). A Tabela 6 apresenta a coleção de parâmetros utilizados no método *Grid Search* para o modelo de geração fotovoltaica.

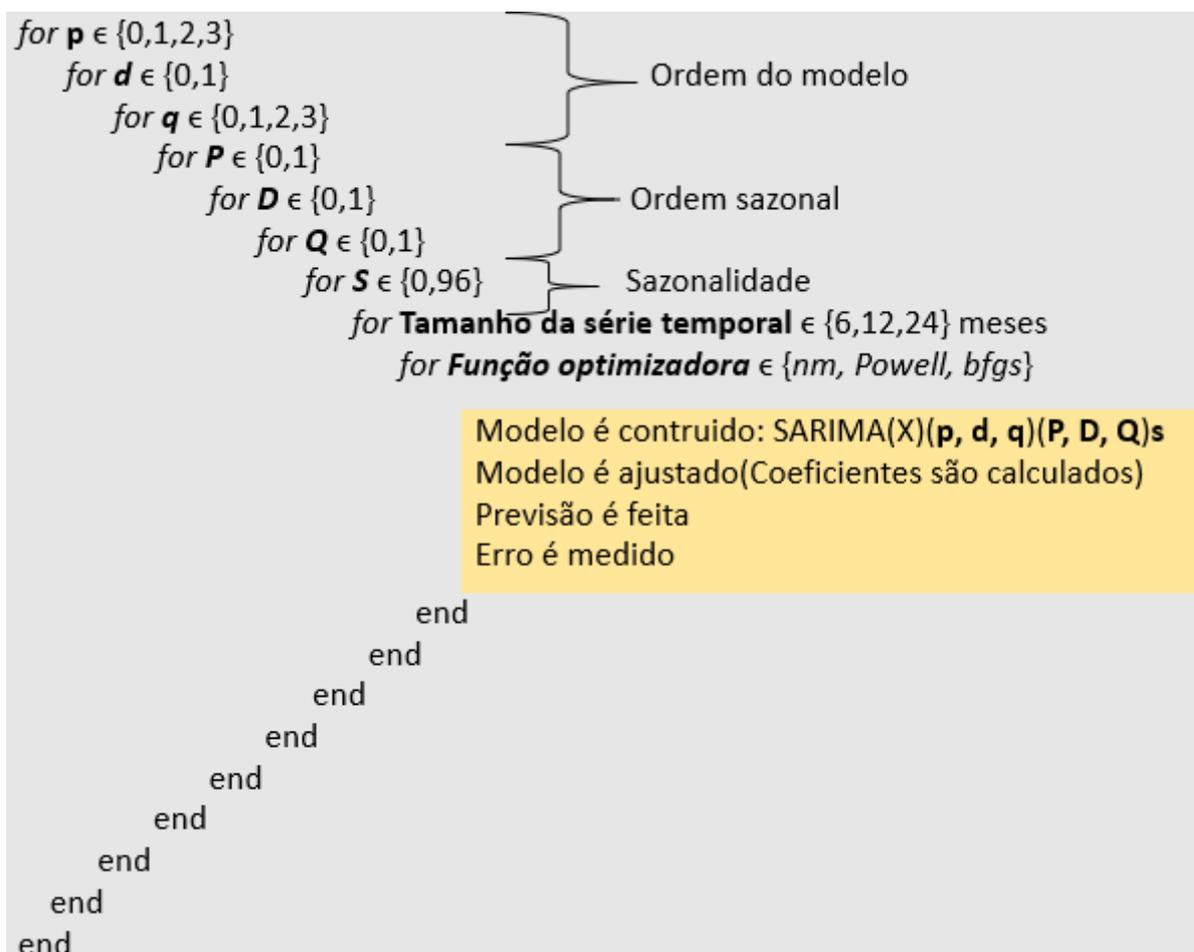
Tabela 6: Opções de parâmetros para *Grid Search* da previsão da geração fotovoltaica

Parâmetros	Símbolos	Possíveis valores
	p	0, 1, 2, 3
Ordem	d	0, 1
	q	0, 1, 2, 3
Ordem sazonal	P	0,1
	D	0,1
	Q	0,1
Sazonalidade	S	96
Método	m	
Tendência	T	'n','c','t','ct'

Com esses parâmetros, é possível construir 3072 modelos diferentes para cada série temporal de geração fotovoltaica. Foram utilizados quatro séries temporais de geração fotovoltaica diferentes e duas maneiras diferentes de expressar a irradiação solar exógena, a irradiação solar exógena original de todas as estações meteorológicas e a irradiação solar média. O número de modelos possíveis então sobe para 27648 modelos diferentes. Quando a sazonalidade é incluída no modelo, o tempo de computação aumenta bastante. Como cada um desses modelos leva em média 25 minutos para ser completado, toda a operação levaria em torno 480 dias em

um computador que não permitisse multiprocessamento. Os parâmetros usados no *Grid Search* são apresentados na Figura 33.

Figura 33: Parâmetros usados no processo de Grid Search em uma forma de laço for.



Com relação ao *Grid Search* no que tange a geração eólica, duas diferentes séries temporais foram utilizadas. A série temporal de geração eólica sem a tendência anual e série temporal de geração eólica sem a tendência anual filtrada. A tendência anual é retirada devido à matriz velocidade de ventos não apresentar esta tendência de crescimento. Com relação aos dados exógenos, utilizou-se a velocidade do vento, a média da velocidade do vento e a velocidade do vento das estações selecionadas com base na correlação com a série temporal de geração eólica. Com os diferentes períodos de tempo para o treinamento do modelo, foi possível criar 6912 modelos. Como este modelo não inclui sazonalidade anual, cada iteração levou no máximo um minuto, exigindo cerca de 5 dias se realizado em computadores que não suportam

multiprocessamento. A Tabela 7 lista os parâmetros usados no *Grid Search* para os conjuntos de dados de geração eólica, carga e carga vertical.

Tabela 7: Parâmetros de entrada Geração eólica, carga e de carga vertical

Parâmetros	Símbolo	Possíveis valores
	p	0, 1, 2, 3
Ordem	d	0, 1
	q	0, 1, 2, 3
Sazonalidade	S	0
Método	m	
Tendência	T	'n','c','t','ct'

A série temporal utilizada no *Grid Search* de carga e carga vertical foram a serie temporal sem tendência anual e a serie temporal sem tendência anual filtrada. Juntamente com os diferentes períodos de tempo para o treinamento dos modelos, foi possível criar 2304 modelos para cada uma das séries temporais.

Cada um desses modelos é executado no Elwetritsch, um centro de computação de alto desempenho que a Universidade de Kaiserslautern compartilha com a Universidade de Mainz. (Alliance for High Performance Computing Rhineland-Palatinate). Para executar esses modelos, é necessário criar um *job*, que é uma lista de requisitos de hardware junto com a tarefa a ser executada. O *script* de tarefa é criado como uma extensão de *shell script*, com arquivos de texto simples contendo linhas com comandos que são executados em sequência pelo interpretador de linha de comando, solicitando o número apropriado de nós, memória, núcleos e o tempo estimado para o trabalho. Se o tempo estimado estiver subdimensionado, o trabalho será eliminado. Por outro lado, se o tempo estimado for superestimado ou os

requisitos de hardware forem muito altos, o trabalho levará muito tempo para ser iniciado até que os requisitos sejam atendidos. A Figura 34 exibe uma parte do *script*.

Figura 34: Trecho do shell script enviado à Elwetritsch

```
#!/bin/bash
##SBATCH -p you_project
#SBATCH -t 72:00:00      # 72 h
#SBATCH --mail-type=END # want a mail notification at end of job
#SBATCH -J GSPV_Amp
#SBATCH -o GSPV_Amp.%j.out
#SBATCH -e GSPV_Amp.%j.err
#SBATCH -N 1           # one node
#SBATCH -n 1           # one task
#SBATCH --cpus-per-task=16
#SBATCH -C HT:off

echo "Working with $SLURM_CPUS_PER_TASK CPU cores"

cd $HOME/Python_Projects/SARIMAX/PVgen_Amp/GS_bfgs_total_2_months/
module add . $HOME/anaconda3.env

python grid_search_SARIMAX.py 2_months bfgs 0 0 0 0 0 0 n PV_std
python grid_search_SARIMAX.py 2_months bfgs 0 0 0 0 0 1 n PV_std
python grid_search_SARIMAX.py 2_months bfgs 0 0 0 0 1 0 n PV_std
python grid_search_SARIMAX.py 2_months bfgs 0 0 0 0 1 1 n PV_std
```

O *Shell script* contém a lista de todos os modelos possíveis com todas as configurações possíveis. O *Python script* é executado através das linhas na parte inferior da **Erro! Fonte de referência não encontrada.** *Python* em azul, instrui o interpretador da linha de comando a vincular o código a seguir, *Grid_seach_SARIMAX.py* ao interpretador *Python*. Este código é comum para todos os modelos e todos os argumentos necessários são passados para o programa externamente. Após a conclusão, todos os arquivos do MATLAB com seus erros e previsões são reunidos em um arquivo e classificados pelo menor erro. Com a coleção dos modelos que produziram os menores erros de previsão, entende-se que os modelos *underfitted* são excluídos desta lista. É então necessário executar um segundo código que separe quais modelos estão *overfitted*, funcionando bem localmente, mas não globalmente, do modelo ideal. Para tanto, a validação cruzada *k-fold* da estrutura é realizada e pode ser dividida em oito seções, conforme Figura 35.

Conjunto de dados e modelo de entrada

Divisão de dados

Construção do modelo

Ajuste do modelo

Walk forward

Previsão

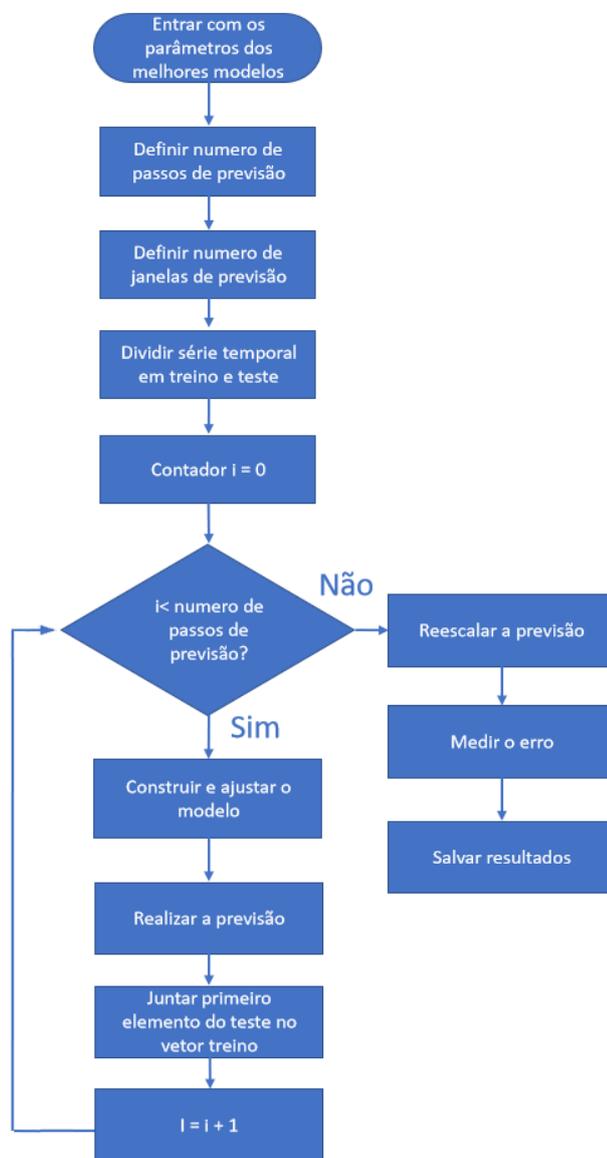
Reescalonamento

Medição de erro

Salvamento dos resultados

Primeiro, a série temporal é dividida de acordo com o comprimento de treinamento; Em seguida, esta série temporal seccionada é usada para o cálculo dos parâmetros do modelo. Como o estágio de validação cruzada tem menos modelos para testar do que o estágio de *Grid Search*, foi possível estender a previsão para 96 passos à frente (um dia). O primeiro item do conjunto de dados de teste é anexado ao conjunto de dados de treinamento, o modelo é novamente equipado com esse novo item e outra previsão é feita - esse procedimento é repetido pela quantidade de vezes definida pelo usuário. Para o conjunto de dados de geração eólica, conjunto de dados de carga e conjunto de dados de carga vertical, foram feitas 200 previsões; para o conjunto de dados geração solar, foram feitas 100 previsões. Após executar a previsão, os erros são medidos e somados, criando uma lista com os parâmetros do modelo, juntamente com a soma do erro.

Figura 35: *Flowchart* da configuração de *walk forward* para achar o melhor modelo



4.3 Modelo de previsão de carga vertical

Prosseguindo com a análise para encontrar o melhor modelo geral de desempenho, o processo de validação cruzada é realizado com base nos melhores modelos de desempenho do estágio do *Grid Search*. A Tabela 8 apresenta uma lista dos melhores modelos de desempenho para previsão de carga vertical.

Tabela 8: Coleção de melhores modelos para conjuntos de dados de carga vertical

Index	Configuration	$\hat{\text{Error}}$	Method	Time
942	('3', '0', '1', '0', '0', '0', 0, 'ct')	104.01780401924945	powell	6_months
951	('3', '0', '1', '0', '0', '0', 0, 't')	104.19841610592844	powell	6_months
618	('3', '1', '2', '0', '0', '0', 0, 'ct')	104.2318611423908	nm	6_months
948	('3', '0', '1', '0', '0', '0', 0, 'n')	104.5176658886097	powell	6_months
945	('3', '0', '1', '0', '0', '0', 0, 'c')	104.58657523634072	powell	6_months
788	('1', '1', '0', '0', '0', '0', 0, 'c')	104.68878055444276	powell	2_year
98	('1', '1', '0', '0', '0', '0', 0, 'c')	104.68879825247788	bfgs	2_year
404	('1', '1', '0', '0', '0', '0', 0, 'c')	104.68879825247788	nm	2_year
100	('1', '1', '0', '0', '0', '0', 0, 'n')	104.70297266510059	bfgs	2_year
407	('1', '1', '0', '0', '0', '0', 0, 'n')	104.70297266510059	nm	2_year
791	('1', '1', '0', '0', '0', '0', 0, 'n')	104.70297271899659	powell	2_year

Após a realização do teste de validação cruzada nos modelos da Tabela 8, o resultado obtido representa a separação do modelo ótimo dos modelos *overfitted*. Nesta circunstância pode-se perceber que o melhor modelo obtido na etapa de *Grid Search*, ou seja, θ_0 , foi na verdade o modelo de pior desempenho na validação cruzada. A lista de validação cruzada é mostrada na Tabela 9

Tabela 9: Resultado da validação cruzada retorna melhor modelo global para carga vertical

Index	Configuration	$\hat{\text{Error}}$	Method	Time
5	('3', '0', '1', '0', '0', '0', 0, 'c')	6832.74	powell	6_months
6	('3', '0', '1', '0', '0', '0', 0, 'n')	6854.14	powell	6_months
7	('3', '0', '1', '0', '0', '0', 0, 't')	7107.06	powell	6_months
0	('1', '1', '0', '0', '0', '0', 0, 'c')	13139.3	bfgs	2_year
1	('1', '1', '0', '0', '0', '0', 0, 'c')	13139.3	nm	2_year
3	('1', '1', '0', '0', '0', '0', 0, 'c')	13139.3	powell	2_year
2	('3', '1', '2', '0', '0', '0', 0, 'ct')	13493.2	nm	6_months
4	('3', '0', '1', '0', '0', '0', 0, 'ct')	196615	powell	6_months

Pode ser visto na Tabela 9 que o erro é muito maior do que na tabela do *Grid Search*. Isso se deve ao fato de que todo erro de cada previsão é somado. Os modelos com melhor desempenho para o conjunto de dados de carga são o SARIMA

o, com 6 meses de dados de treinamento e usando o algoritmo de Powell para calcular os coeficientes.

4.4 Modelo de previsão de carga

Os modelos de melhor desempenho obtidos através da etapa de *Grid Search* para o conjunto de dados de carga são mostrados na Tabela 10.

Tabela 10: Modelos de previsão de carga retornados da etapa *Grid Search* com menores erros

Index	Configuration	$\hat{\text{Error}}$	Method	Time
517	('1', '0', '3', '0', '0', '0', 0, 'ct')	244.1388056627868	nm	2_year
743	('3', '1', '1', '0', '0', '0', 0, 't')	245.71577319857874	nm	6_months
671	('2', '1', '3', '0', '0', '0', 0, 't')	246.44342130164537	nm	6_months
278	('2', '1', '3', '0', '0', '0', 0, 'ct')	246.63018586387358	bfgs	6_months
276	('2', '1', '3', '0', '0', '0', 0, 'ct')	246.64162777663154	bfgs	1_year
372	('3', '1', '3', '0', '0', '0', 0, 'ct')	246.65088355658847	bfgs	1_year
285	('2', '1', '3', '0', '0', '0', 0, 't')	246.84925518066987	bfgs	1_year
662	('2', '1', '3', '0', '0', '0', 0, 'ct')	246.87171684328834	nm	6_months
757	('3', '1', '3', '0', '0', '0', 0, 'ct')	246.8868652364384	nm	2_year

Com os melhores modelos para previsão de carga, é necessário, então, discriminar os modelos *overfitted* do modelo ótimo. Depois de realizar o processo de validação cruzada obtém-se a Tabela 11

Tabela 11: Resultados da validação cruzada para o conjunto de dados de carga

Index	Configuration	$\hat{\text{Error}}$	Method	Time
4	('1', '0', '3', '0', '0', '0', 0, 'ct')	123345	nm	2_year
0	('2', '1', '3', '0', '0', '0', 0, 'ct')	138885	bfgs	1_year
1	('2', '1', '3', '0', '0', '0', 0, 'ct')	139270	bfgs	2_year
8	('3', '1', '3', '0', '0', '0', 0, 'ct')	139687	nm	2_year
2	('2', '1', '3', '0', '0', '0', 0, 'ct')	140579	bfgs	6_months
5	('2', '1', '3', '0', '0', '0', 0, 'ct')	143336	nm	6_months
7	('3', '1', '1', '0', '0', '0', 0, 't')	144636	nm	6_months
6	('2', '1', '3', '0', '0', '0', 0, 't')	144914	nm	6_months
3	('3', '1', '3', '0', '0', '0', 0, 'ct')	146824	bfgs	1_year

O melhor modelo para previsão de carga foi o SARIMA 0
treinado com 2 anos de dados passados e usando o algoritmo *Nelder-Mead* para
calcular os coeficientes.

4.5 Modelo de previsão de geração fotovoltaica

O resultado do *Grid Search* para geração fotovoltaica pode ser visto na Tabela 12, onde estão listados os melhores modelos para previsão fotovoltaica.

Tabela 12: Melhores modelos para previsão fotovoltaica obtidos a partir do processo de *Grid Search*

Index	Configuration	Error	Method	Time	Dataset	Error_scaled
4574	('1', '1', '3', '0', '0', '1', 96, 'n')	34.18523270127265	nm	6_months	Origin	34.185232701...
5899	('3', '0', '3', '1', '1', '1', 96, 't')	0.010923504489799559	nm	6_months	PV_max	48.1885
4025	('1', '0', '1', '1', '1', '1', 96, 'n')	0.014278322002723902	nm	6_months	PV_max	62.9848
4549	('1', '1', '3', '0', '0', '0', 96, 'ct')	0.01443825519198155	nm	1_year	PV_max	63.695
4021	('1', '0', '1', '1', '1', '1', 96, 'ct')	0.014729988739438125	nm	6_months	PV_max	64.979
4123	('1', '0', '2', '1', '1', '0', 96, 't')	0.01502164880172501	nm	6_months	PV_max	66.266
4119	('1', '0', '2', '1', '1', '0', 96, 'c')	0.015028161466744428	nm	6_months	PV_max	66.2947
1451	('1', '1', '3', '0', '0', '0', 96, 'c')	0.015182648780727335	bfgs	1_year	PV_max	66.9806

Usando os modelos do processo de *Grid Search*, a validação cruzada retornou o modelo de melhor desempenho geral. O resultado da validação cruzada pode ser visto em Tabela 13

Tabela 13: Resultados de validação cruzada para geração de PV

Index	Configuration	Error	Method	Time	Dataset
1	('1', '0', '2', '1', '1', '0', 0, 'c')	1480.97	nm	6_months	PV_max
6	('3', '0', '3', '1', '1', '1', 0, 't')	1549.31	nm	6_months	PV_max
0	('1', '0', '1', '1', '1', '1', 0, 'n')	2649.6	nm	6_months	PV_max
5	('1', '1', '3', '0', '0', '1', 0, 'n')	1.0734e+09	nm	6_months	Origin

Pode-se observar que o modelo de melhor desempenho é o SARIMAX₉₆ com 6 meses de dados de treinamento e usando o algoritmo Nelder-Mead para calcular os coeficientes.

4.6 Modelo de previsão de geração eólica

Com respeito aos modelos de previsão de energia eólica a Tabela 14 mostra os modelos com menores erros de acordo com o processo de *Grid Search*

Tabela 14: Coleção dos modelos de previsão de geração eólica oriundo do processo de *Grid Search*

Index	Configuration	$\hat{\text{Error}}$	Method	Time
571	('2', '0', '2', '0', '0', '0', 0, 'c')	69.0242353889137	nm	1_year
576	('2', '0', '2', '0', '0', '0', 0, 'n')	69.72912280058385	nm	6_months
654	('3', '0', '1', '0', '0', '0', 0, 'ct')	70.17781646969036	nm	6_months
657	('3', '0', '1', '0', '0', '0', 0, 'c')	70.33645825878982	nm	6_months
660	('3', '0', '1', '0', '0', '0', 0, 'n')	70.52106069173142	nm	6_months
568	('2', '0', '2', '0', '0', '0', 0, 'ct')	70.53264926474816	nm	1_year
574	('2', '0', '2', '0', '0', '0', 0, 'n')	71.20607825120581	nm	1_year
663	('3', '0', '1', '0', '0', '0', 0, 't')	71.35544113031798	nm	6_months

O processo de validação cruzada é então realizado para separar o modelo ótimo dos modelos *overfitted*. O resultado da validação cruzada é mostrado na Tabela 15.

Tabela 15: Resultados da validação cruzada dos modelos de geração eólica

Index	Configuration	Error	Method	Time
3	('2', '0', '2', '0', '0', '0', 0, 'c')	85088.4	nm	6_months
11	('3', '0', '1', '0', '0', '0', 0, 't')	85541.5	nm	6_months
7	('3', '0', '1', '0', '0', '0', 0, 'ct')	85604.2	nm	6_months
9	('3', '0', '1', '0', '0', '0', 0, 'c')	88740.6	nm	6_months
10	('3', '0', '1', '0', '0', '0', 0, 'n')	91405.3	nm	6_months
0	('2', '0', '1', '0', '0', '0', 0, 'n')	92390	nm	6_months
5	('2', '0', '2', '0', '0', '0', 0, 't')	113100	nm	1_year
1	('2', '0', '2', '0', '0', '0', 0, 'ct')	114892	nm	1_year
2	('2', '0', '2', '0', '0', '0', 0, 'c')	118325	nm	1_year
6	('3', '0', '1', '0', '0', '0', 0, 'ct')	123375	nm	1_year

Pode ser visto a partir da Tabela 15 O melhor modelo obtido na validação cruzada é diferente do melhor modelo obtido no processo de *Grid Search*. O modelo de melhor desempenho é o 0 com 6 meses de dados de treinamento e algoritmo de *Nelder-Mead* para calcular os coeficientes.

4.7 Conclusão

Comparando o tempo de computação esperado em um computador sem múltiplos núcleos, é visível a importância do computador de alto desempenho Elwetritsch. Uma simulação que exigiria um ano e meio de poder computacional constante foi encurtada para cerca de um mês de simulação. Avaliando os resultados do *Grid Search* e da validação cruzada, particularmente no caso de carga vertical, é claro que o modelo de melhor desempenho no *Grid Search* foi o modelo de pior desempenho na validação cruzada. Isso nos informa que esse modelo provavelmente é o modelo *overfitted*. Outra conclusão é que os possíveis parâmetros *Grid Search* foram escolhidos com cuidado, evitando o caso de produzir muitos modelos *overfitted*. A combinação de *Grid Search* e validação cruzada foi necessária para poder discriminar apenas um modelo da lista inteira gerada no processo de *Grid Search*. A Tabela 16 resume os melhores modelos obtidos a partir do processo de validação cruzada com seus respectivos coeficientes.

Tabela 16: Resumo dos resultados obtidos no processo de validação cruzada.

Série temporal	Modelo	Parâmetros do modelo
Geração solar	SARIMAX(1,0,2)x(1,1,0) ₉₆	
Geração eólica	ARMAX(2,2)	
Carga	ARMA(1,3)	
Carga vertical	ARMA(3,1)	

5 Resultados de previsão

5.1 Introdução

Neste capítulo, a previsão de cada conjunto de dados é realizada usando os melhores modelos obtidos com o processo de *Grid Search* e validação cruzada. A previsão é feita com um horizonte de previsão de 8 horas à frente e 1 dia à frente. O erro de previsão é medido com a métrica de erro RMSE e com o erro de porcentagem média. É realizada uma medição de erro média relacionada ao afastamento da previsão a partir do tempo real.

5.2 Resultados da previsão de carga

5.2.1 Múltiplas previsões de 8 horas

Para recuperar informações estatísticas do desempenho do modelo, são feitas 200 previsões de carga. As previsões foram compostas por uma previsão de um único passo de 8 horas, feita com o modelo ARMA (1,3) estabelecido com o processo de *Grid Search* e validação cruzada. A Figura 36 mostra a previsão múltipla de 8 horas para o futuro, composta de 32 valores cada, variando de 29/07/2016 das 7:00 às 31/07/2016 às 16:45 mostradas em azul. A curva laranja corresponde ao conjunto de dados de carga original.

O erro RMSE da previsão múltipla de 8 horas é calculado para essas 200 previsões. Após, a média do RMSE é calculada a partir do final dos dados de treinamento. A Figura 37 mostra o aumento médio do erro RMSE de acordo com o quão longe a previsão se encontra do tempo real.

A Figura 38 mostra a média do erro percentual, com o afastar da previsão do tempo real. Isso é útil para mostrar como o erro está relacionado à magnitude da carga.

Figura 36: 200 previsões de carga de 8 horas

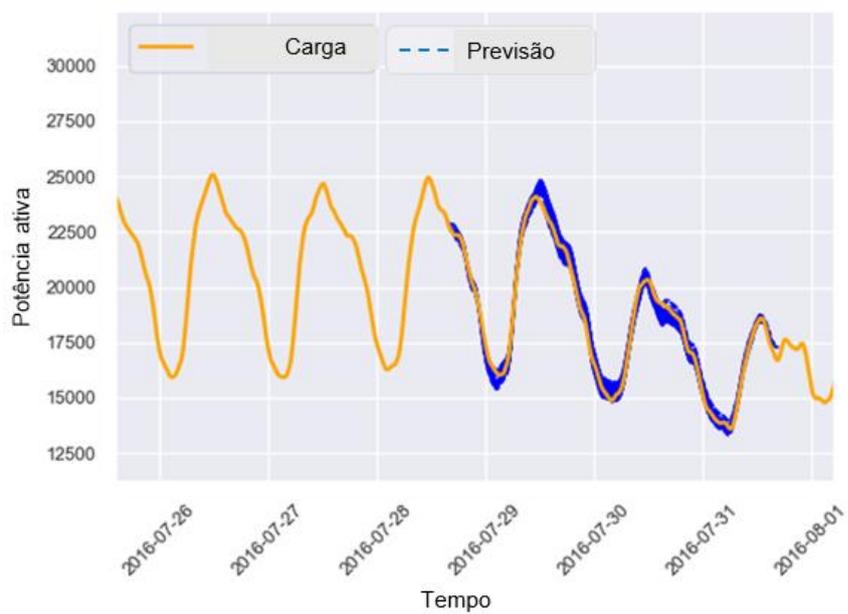


Figura 37: Crescimento do erro RMSE para o conjunto de dados de carga com o aumento do horizonte de previsão até no máximo 8 horas

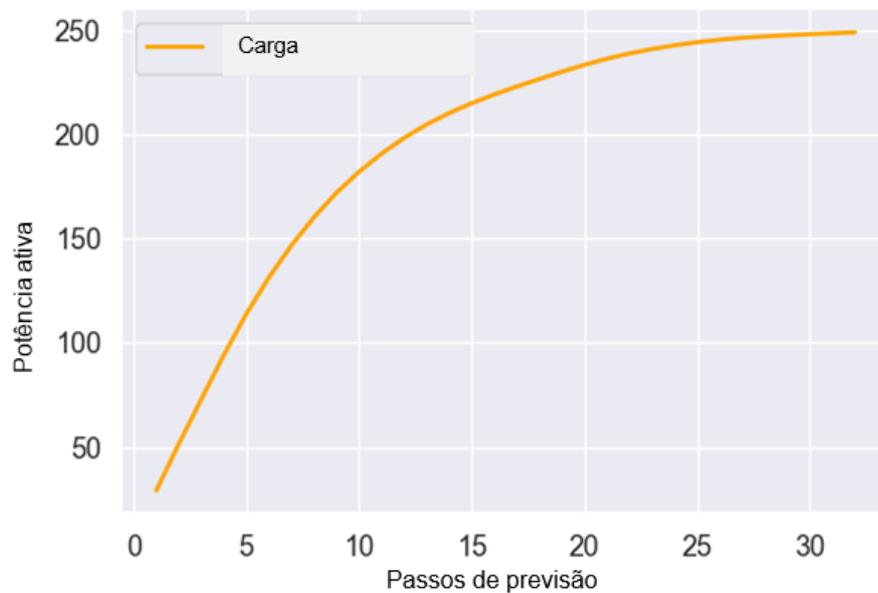
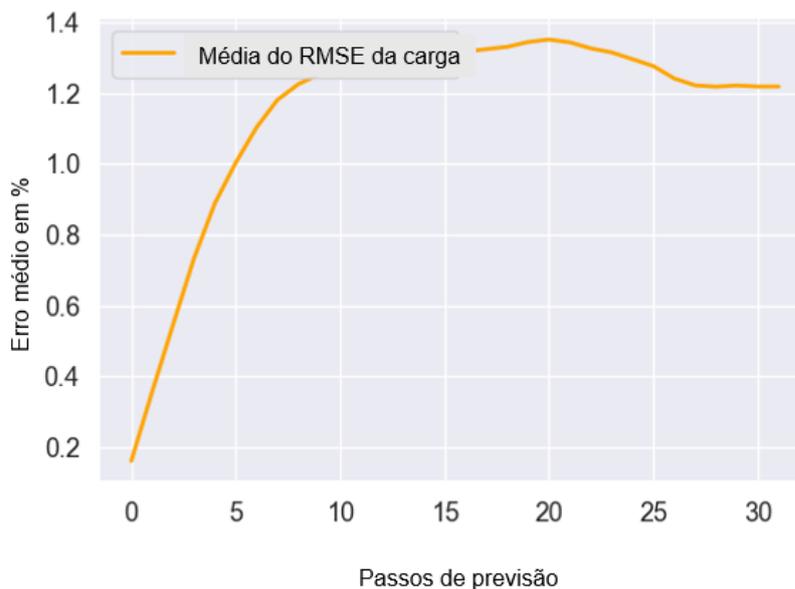


Figura 38: Crescimento do erro percentual médio de carga com o aumento do horizonte de previsão até no máximo 8 horas



5.2.2 Múltiplas previsões de 24 horas pro futuro

A previsão de 24 horas, composta por uma única previsão de 96 valores para o conjunto de dados de carga é mostrada na Figura 39. A curva múltipla azul representa a previsão de carga e a curva laranja representa o conjunto de dados de carga original; nesta figura existem 200 previsões.

Seguindo o mesmo procedimento da previsão de oito horas em várias etapas, o RMSE é calculado e é calculada a média do RMSE a medida que a previsão se afasta do tempo real, conforme mostrado na Figura 40.

O gráfico do erro percentual médio em relação à distância de previsão do conjunto de dados de treinamento de carga é mostrado **Erro! Fonte de referência não encontrada..**

Figura 39: 200 previsões de carga de 24 horas para o futuro

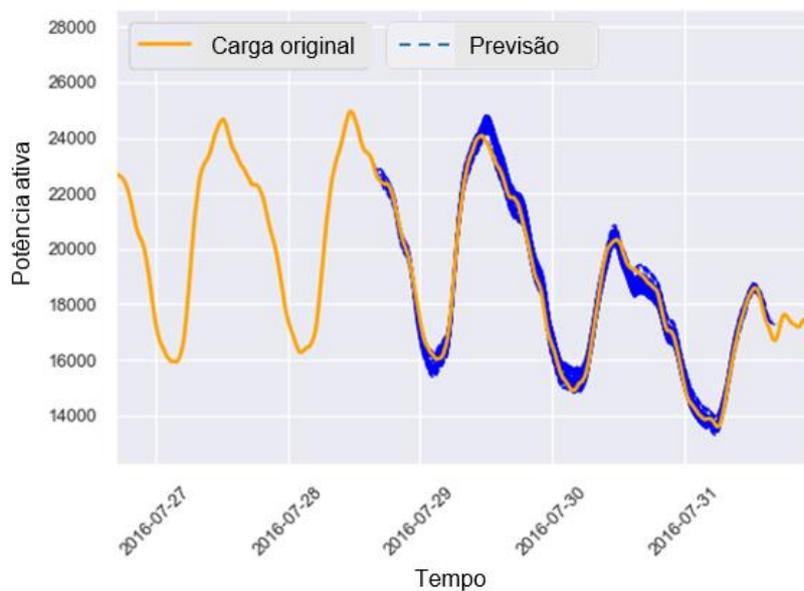


Figura 40: Crescimento do erro RMSE para o conjunto de dados de carga com o aumento do horizonte de previsão até no máximo 24 horas

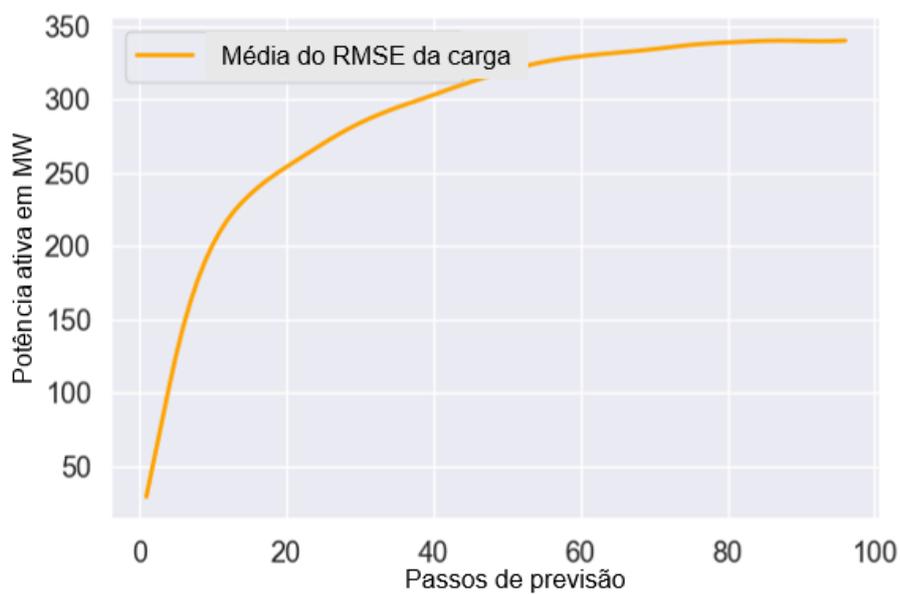
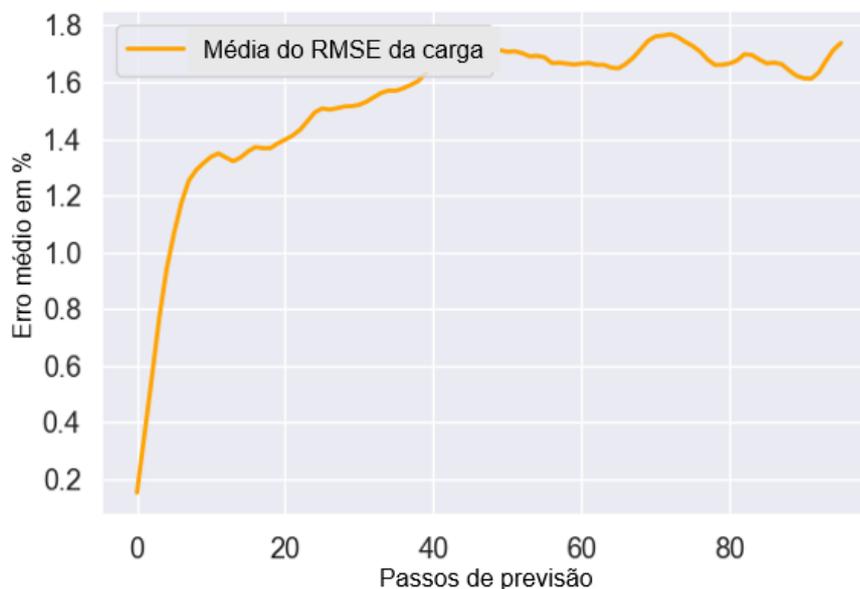


Figura 1: Crescimento do erro percentual médio de carga com o aumento do horizonte de previsão até no máximo 24 horas



5.3 Resultados da previsão de carga vertical

5.3.1 Múltiplas previsões de 8 horas

Figura 42 ilustra a previsão múltipla de oito horas de carga vertical, composta de 32 valores de previsão cada, variando de 29/07/2016 às 7:00 às 31/07/2016 às 16:45 mostradas em azul., neste caso, 200 previsões foram realizadas. A curva laranja representa o conjunto de dados de carga verticais original.

O erro RMSE entre as previsões e os dados originais é então computado, a medida que a previsão se afasta do tempo real. O resultado de 200 previsões é mostrado em Figura 43

Figura 44 apresenta o erro percentual médio em relação ao aumento do horizonte temporal de previsão, sendo o horizonte máximo 8 horas.

Figura 41: 8 horas de previsão da carga vertical

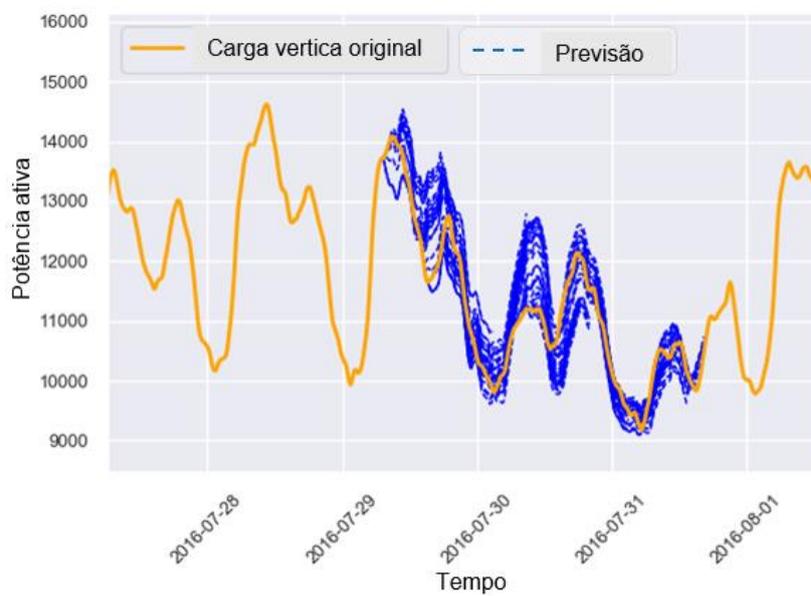


Figura 42: Crescimento do erro de RMSE para o conjunto de dados de carga vertical com o aumento do horizonte de previsão até no máximo 8 horas

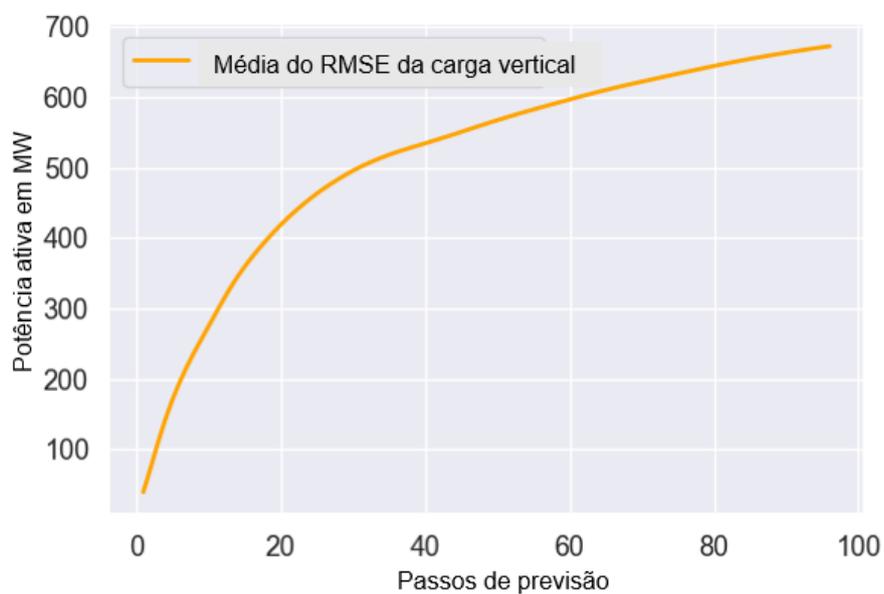
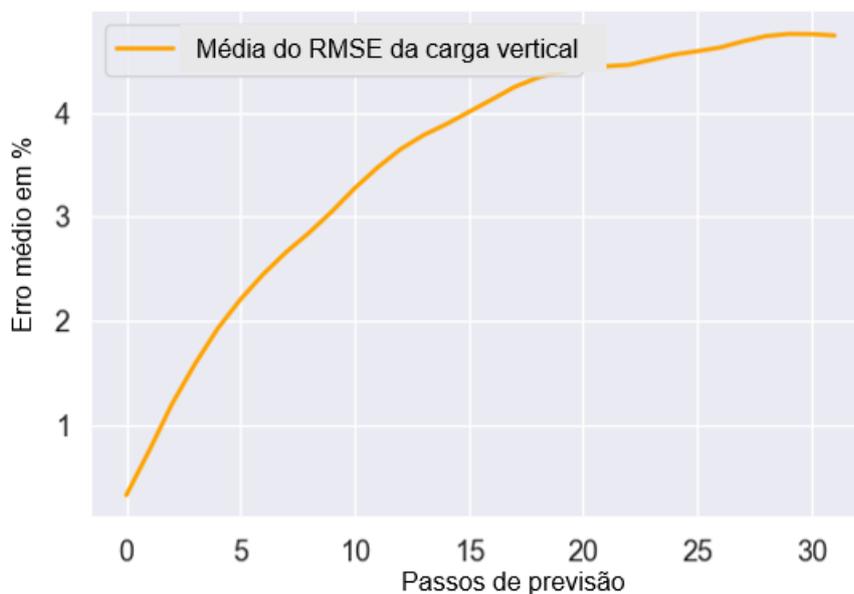


Figura 43: Crescimento do erro percentual médio da carga vertical com o aumento do horizonte de previsão até no máximo 8 horas



5.3.1 Múltiplas previsões de 24 horas pro futuro

Figura 44 apresenta a previsão de múltiplas etapas de carga vertical de 24 horas feita com ARMA (3,1, 'c') junto com o conjunto de dados de carga vertical original. Figura 45 representa o erro RMSE entre as previsões e os dados originais com relação com o aumento do horizonte de previsão. Figura 46 apresenta o erro de porcentagem média com o aumento do horizonte de previsão em relação ao tempo real.

Figura 44: Múltiplas previsões de 24 horas da carga vertical junto com a carga vertical original

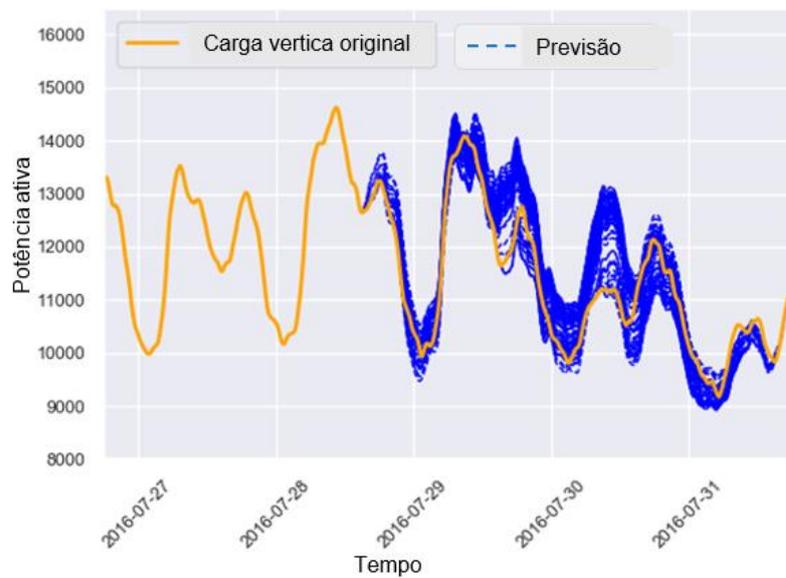


Figura 45: Crescimento do erro RMSE para o conjunto de dados de carga vertical com o aumento do horizonte de previsão até no máximo 24 horas a partir do tempo real

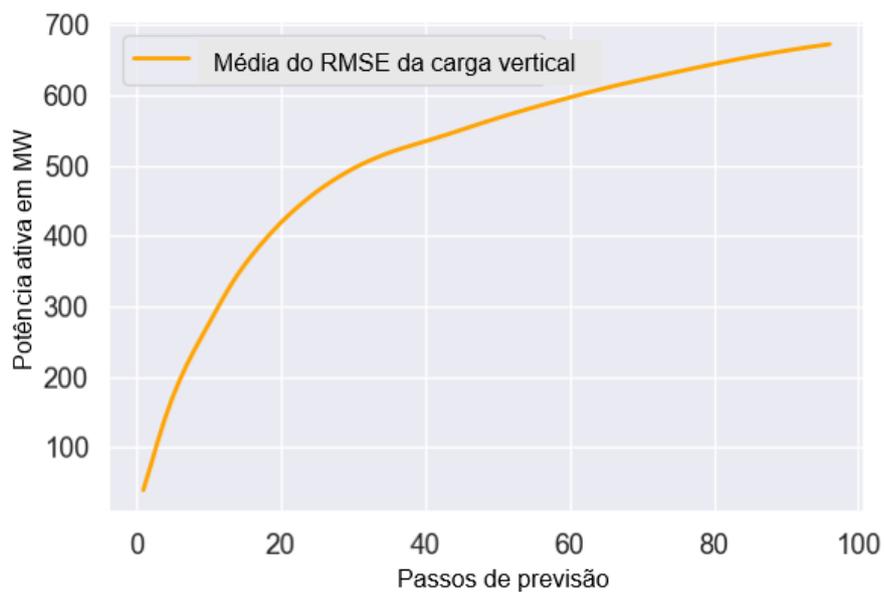
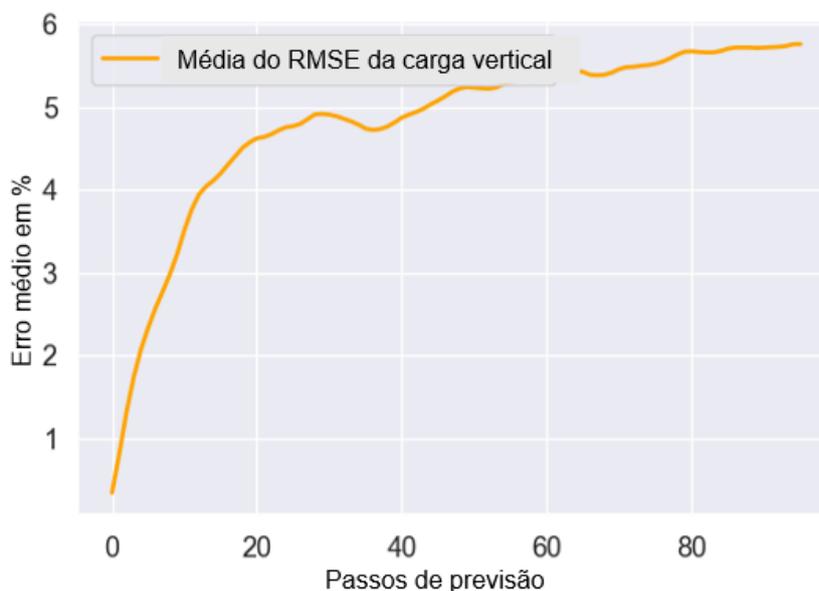


Figura 46: Crescimento do erro percentual médio da carga vertical com o aumento do horizonte de previsão até no máximo 24 horas a partir do tempo real



5.4 Resultados da previsão de geração fotovoltaica

5.4.1 Múltiplas previsões de 8 horas pro futuro

Uma vez que modelos que abordem sazonalidade requerem muito mais poder computacional, exigindo assim muito mais tempo, no processo de validação cruzada, foram feitas 100 previsões, ao invés de 200 como nos casos anteriores. O modelo que se sobressaiu foi o SARIMAX com o aumento do horizonte de previsão até no máximo 8 horas a partir do tempo real

Figura 47 em seguida, apresenta múltiplas previsões de oito horas sobreposta à série temporal de geração fotovoltaica original. Um vetor binário determinístico é usado para corrigir a geração fotovoltaica, produzindo 0 MW quando existe a certeza de não haver sol.

Figura 48 apresenta o crescimento médio do erro RMSE com o aumento do horizonte de previsão até no máximo 8 horas a partir do tempo real

Figura 49 representa o gráfico do erro percentual médio com o aumento do horizonte de previsão até no máximo 8 horas a partir do tempo real

Figura 47: 100 múltiplas previsões de 8 horas para a geração fotovoltaica com o conjunto original

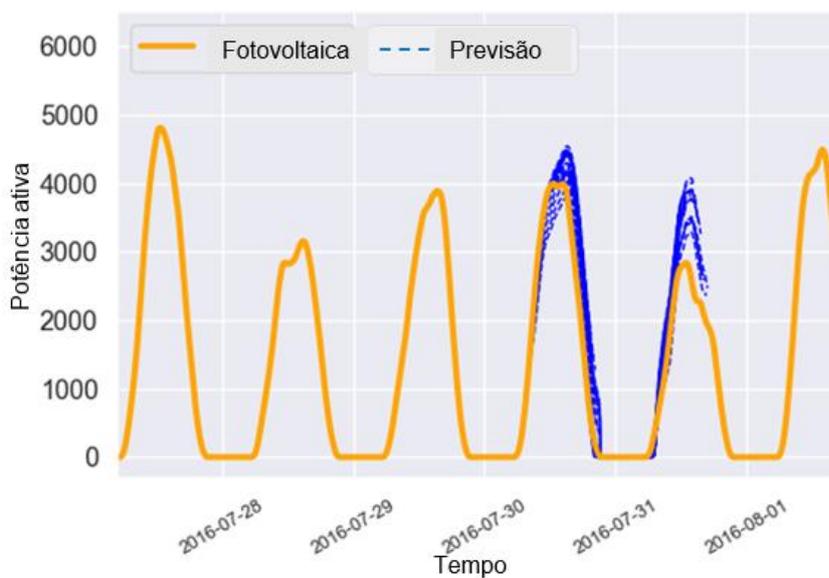


Figura 48: Crescimento do erro RMSE para a geração fotovoltaica com o aumento do horizonte de previsão até no máximo 8 horas a partir do tempo real

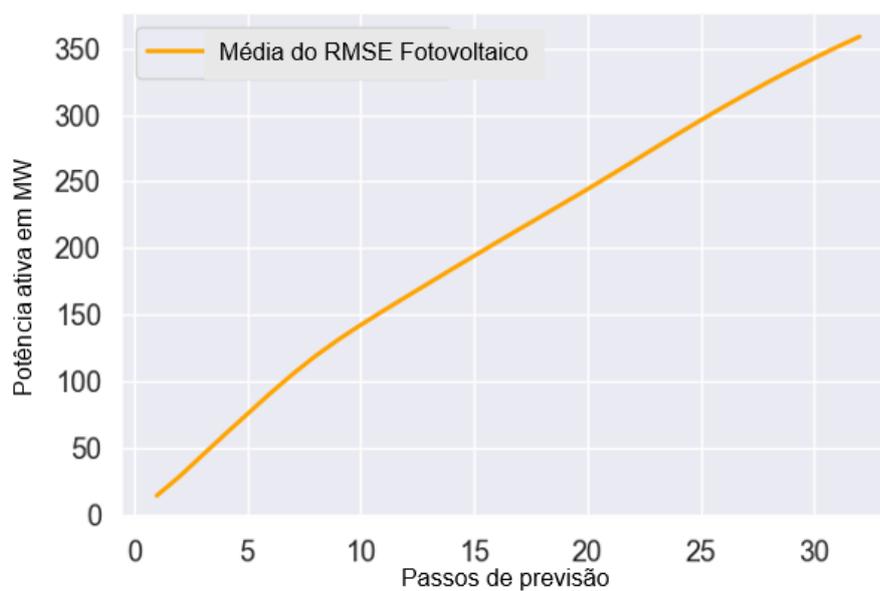
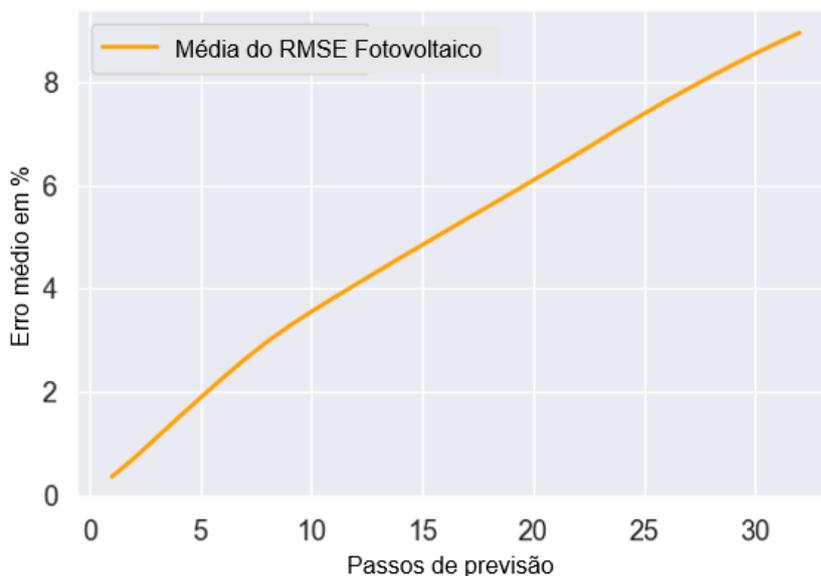


Figura 49: Crescimento do erro médio percentual para a geração fotovoltaica com o aumento do horizonte de previsão até no máximo 8 horas a partir do tempo real



5.4.1 Múltiplas previsões de 24 horas pro futuro

com o aumento do horizonte de previsão a partir do tempo real.

Figura 50 apresenta múltiplas previsões de geração fotovoltaica de 24 horas juntamente com o conjunto de dados fotovoltaicos original. Aqui, pode-se ver a influência do vetor binário perto dos momentos em que o sol nasce ou quando o sol se põe.

Figura 51 retrata o crescimento médio do erro RMSE com o aumento do horizonte de previsão até no máximo 24 horas a partir do tempo real

Seguindo o mesmo padrão de representação de dados, Figura 52 exhibe a previsão do erro percentual médio a partir de 24 horas com o aumento do horizonte de previsão a partir do tempo real.

Figura 50: Múltiplas previsões de 24 horas para o futuro para a geração fotovoltaica

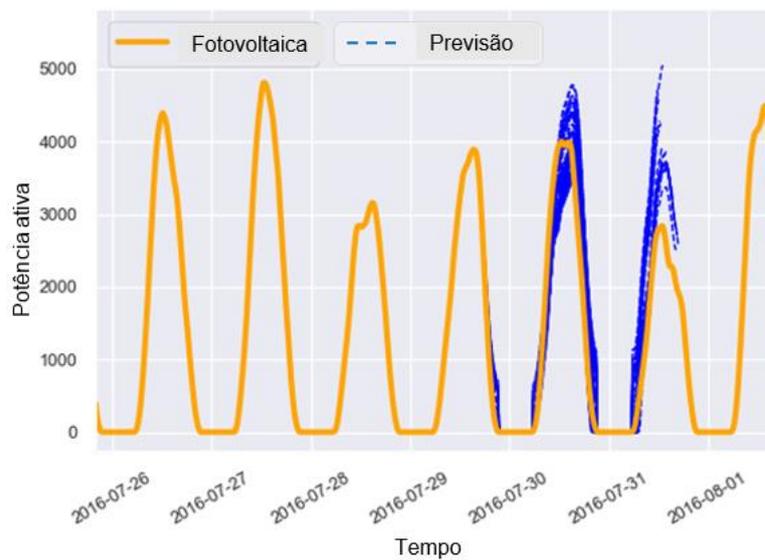


Figura 51: Crescimento do erro RMSE das previsões de geração fotovoltaica com o aumento do horizonte de previsão de até 24 horas

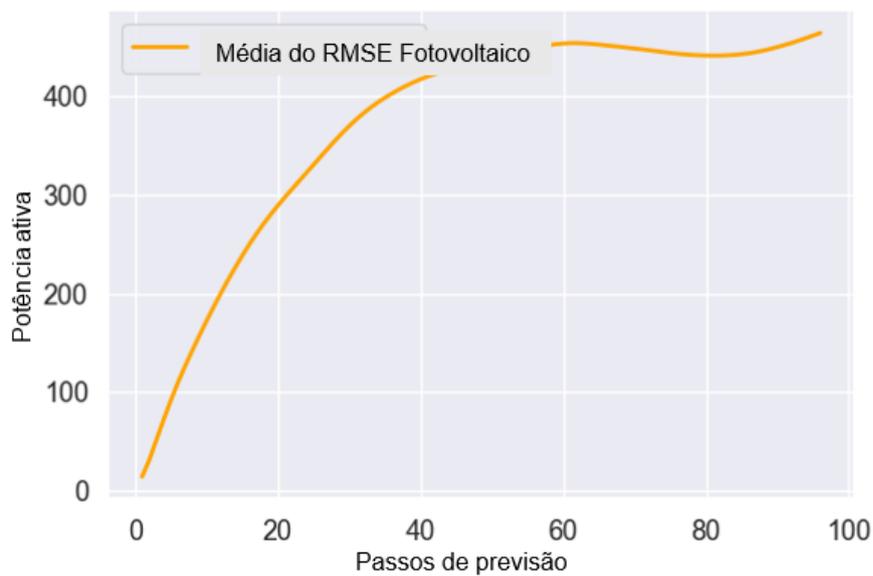
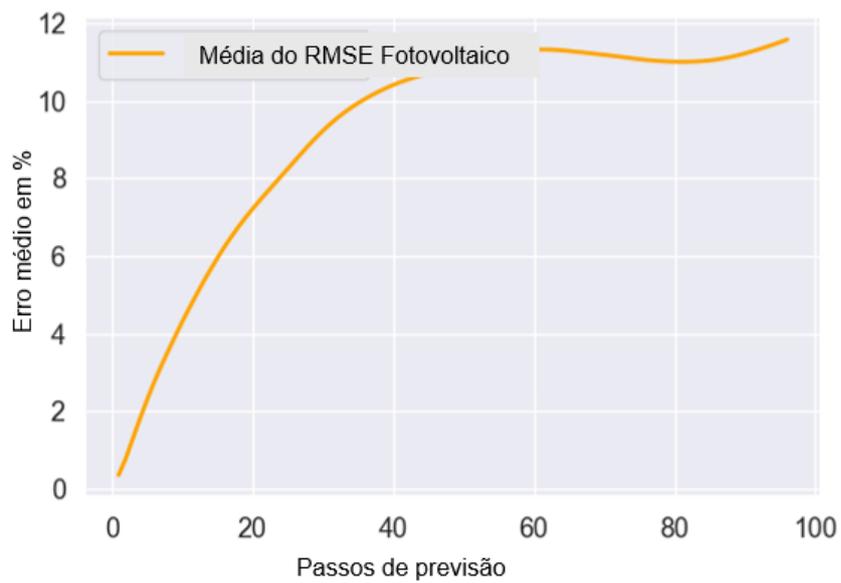


Figura 52: Crescimento do erro percentual médio das previsões de geração fotovoltaica com o aumento do horizonte de previsão de até 24 horas



5.5 Resultados da previsão de geração de energia eólica

5.5.1 Múltiplas previsões de 8 horas

As múltiplas previsões de geração eólica de 8 horas foram feitas usando o ARMAX. As previsões são apresentadas na Figura 53 junto com o conjunto de dados de geração eólica original.

Figura 54 apresenta o crescimento médio do erro RMSE com o aumento do horizonte de previsão até no máximo 8 horas a partir do tempo real para 200 amostras

Figura 53: 200 múltiplas previsões de geração eólica



Figura 55 exibe o gráfico do erro percentual médio das previsões de geração eólica com o aumento do horizonte de previsão até no máximo 8 horas a partir do tempo real.

Figura 54: Crescimento do erro RMSE das previsões de geração eólica com o aumento do horizonte de previsão de até 8 horas

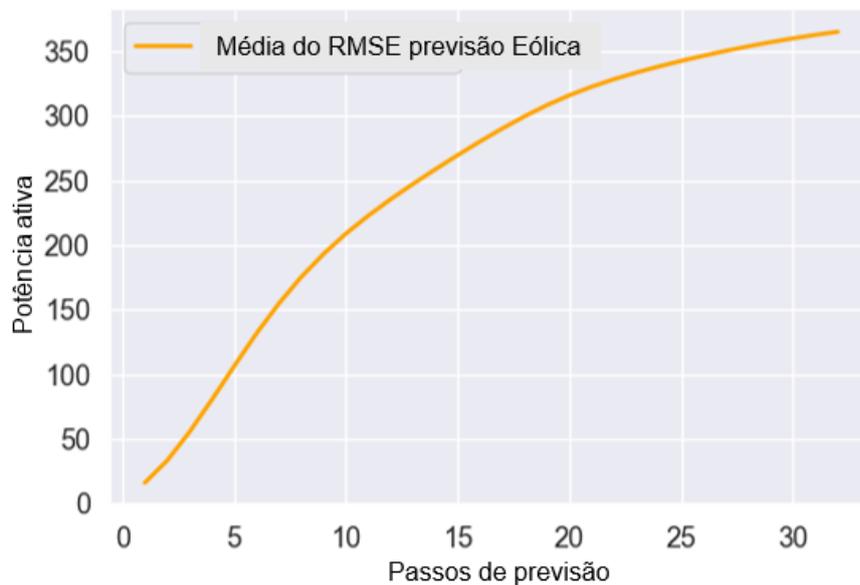
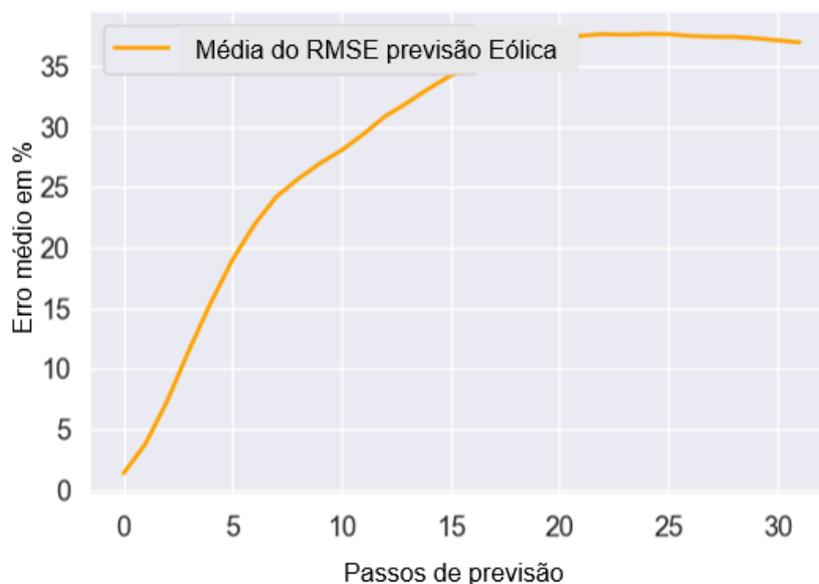


Figura 55: Crescimento do erro percentual médio das previsões de geração eólica com o aumento do horizonte de previsão de até 8 horas



5.5.1 Múltiplas previsões de 24 horas pro futuro

para as múltiplas previsões de até 24 horas de geração eólica

Figura 56 apresenta múltiplas previsões de 24 horas de geração eólica em conjunto com a geração eólica original.

Figura 57, apresenta o crescimento médio do erro RMSE com o aumento do horizonte de previsão em até 24 horas

Figura 58 mostra a previsão do erro percentual médio para as múltiplas previsões de até 24 horas de geração eólica

Figura 56: 200 múltiplas previsões de 24 horas de geração eólica



Figura 57: Crescimento do erro RMSE para previsões de 24 horas de geração eólica

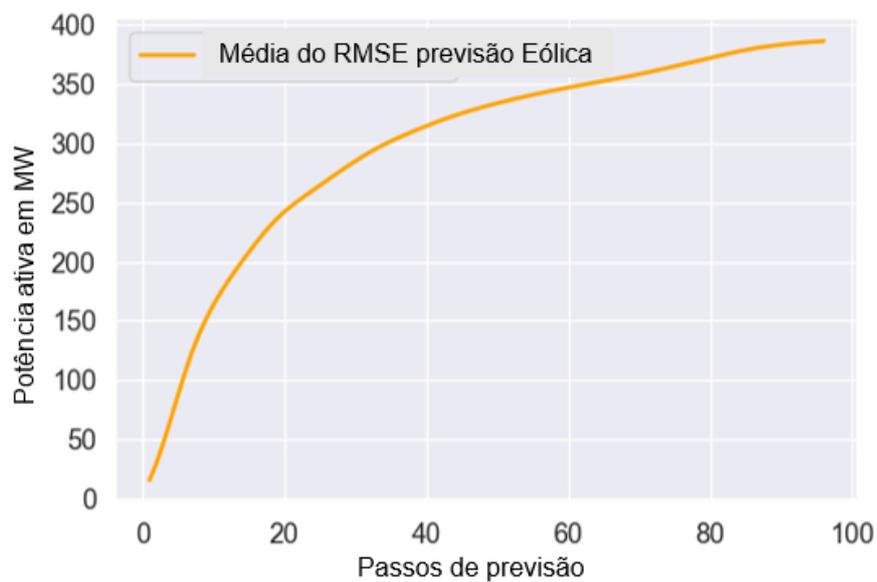
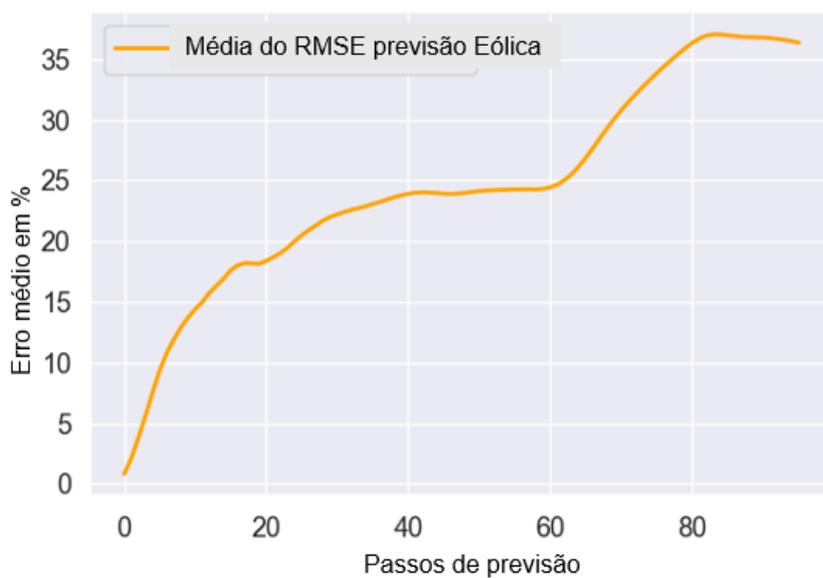


Figura 58: Crescimento do erro percentual médio para previsões de 24 horas de geração eólica



6 Conclusão

O presente trabalho é motivado segundo a afirmativa de que o aumento das incertezas das condições da rede, trazidas pelo crescimento de fontes renováveis, vê em previsões estatísticas, uma forma de definir um possível e provável estado da rede. A criação de modelos de previsão de geração fotovoltaica, geração eólica, carga e carga vertical utilizando para isso dados históricos criando assim uma ferramenta para uso na operação do sistema de potência.

A previsão de carga apresentou o melhor desempenho dentre as variáveis de previsão. O erro permaneceu relativamente constante entre 1,4% e 1,8% entre 5 horas e 24 horas. Para uma previsão com menos de 5 horas, o erro foi ainda menor. Outra previsão de bom desempenho foi a carga vertical, que também mostrou um erro de previsão relativamente constante com o afastar do tempo real. O erro percentual para a carga vertical variou entre 4% e 6% entre 5 e 24 horas. Tanto a série temporal de carga como a carga vertical, após a manipulação, apresentavam características bem próximas a estacionariedade.

Quanto ao modelo de previsão de geração fotovoltaica, o erro de previsão percentual apresentou um aumento constante com o afastar do tempo real. Também foi necessário usar um vetor binário para corrigir a previsão de geração fotovoltaica em horários onde era impossível ter geração fotovoltaica. Modelos estatísticos não são os mais recomendados para a previsão de geração fotovoltaica pelo fato desta série temporal apresentar não linearidades.

Referente à geração eólica, o erro percentual tende a se estabilizar em algum momento, visível através da aproximação de uma linha reta no gráfico de erro percentual, porém, após alguns períodos, o erro aumenta novamente. Esse comportamento ocorre devido à diminuição da magnitude da geração eólica original, que na fórmula da porcentagem é o denominador. Assim, um denominador menor mostra um aumento no erro percentual. O modelo de previsão de geração eólica

poderia ser melhorado utilizando um modelo estatístico que pode abordar a volatilidade na série temporal, como os modelos GARCH ou EGARCH.

Uma desvantagem dos modelos estatísticos é o fato de que não levam em consideração a ocorrência de feriados para a previsão de carga e carga vertical. Os métodos estatísticos de previsão não são apropriados para modelar a presença de feriados devido a não linearidade. Uma maneira possível de contornar este problema é através do uso de métodos não lineares tais como métodos de *machine learning*.

Ao avaliar os resultados da previsão, fica clara a importância de aproximar os conjuntos de dados das características de estacionariedade, uma vez que as previsões de melhor desempenho tinham os conjuntos de dados de entrada para os modelos com características mais próximas dos requisitos de estacionariedade.

REFERENCIAS BIBLIOGRAFICAS:

(BMWI), F. M. F. E. A. A. E. **An electricity market for Germany's energy transition.** [S.l.]. 2015.

Unsupported source type (Interview) for source Ste18.

Unsupported source type (Interview) for source Placeholder1.

AGORA ENERGIEWENDE. **Projected EEG Costs up to 2035.** [S.l.]: [s.n.], 2016.

Alliance for High Performance Computing Rhineland-Palatinate. Disponível em: <<https://www.ahrp.info/>>.

AMPRION. Amprion, fev. 2019. Disponível em: <www.amprion.net>.

BASHIRAHAMAD FARDIN MOMIN, G. C. **Univariate Time Series Models for Forecasting Stationary and Non-stationary Data: A Brief Review.** [S.l.]. 2017.

BOVAS ABRAHAM, J. L. **Statistical Methods for Forecasting.** [S.l.]: [s.n.], 1983.

BROCKWELL, P. J. A. D. R. A. **Introduction to Time Series and.** New York: [s.n.], 1996.

BRUNNER, C. **Changes in electricity spot price formation in Germany caused by a high share of renewable energies.** [S.l.]: [s.n.], 2014.

BRUNNO BURGER. <https://www.energy-charts.de>, 07 fev. 2019. Disponível em: <https://www.energy-charts.de/ren_share_de.htm?source=ren-share&period=annual&year=all>. Acesso em: 05 mar. 2019.

CHAIR OF STATISTICS, UNIVERSITY OF WÜRZBURG. **A First Course on Time Series Analysis.** [S.l.]: [s.n.], 2012.

CHRISTIAN GROWITSCH, R. N. **Electricity spot trading in Germany: Price formation and convergence**. [S.I.]. 2011.

DEUTSCHER Wetterdienst. **Deutscher Wetterdienst**. Disponivel em: <<https://www.dwd.de/DE/leistungen/klimadatenweltweit/stationsverzeichnis.html?lsblid=374532>>.

ENERGIEBILANZEN. Disponivel em: <<https://www.ag-energiebilanzen.de/>>.

ENTSOE. **entsoe**. Disponivel em: <<https://transparency.entsoe.eu/>>.

IFEU. Disponivel em: <<https://www.ifeu.de/en/project/renewable-energy-sources-act-eeg-progress-report/>>.

JOHN Y. CAMPBELL, A. W. L. & A. C. M. **The Econometrics of Financial Markets**. [S.I.]. 1997.

KÜHNERT, H. J. **Development of a photovoltaic power prediction system for forecast horizons of several hours**. [S.I.]. 2016.

MÖLLER, C. **Balancing energy in the German market design**. [S.I.]. 2010.

RICARDO BESSA, C. M. B. S. M. M. Handling renewable energy variability and uncertainty in power systems operation, 2014.

RICARDO BESSA, C. M. S. M. **Handling renewable energy variability and uncertainty in power systems operation**. [S.I.]. 2013.

TSAY, R. S. **Analysis of financial time series**. [S.I.]: [s.n.], 2010.

VAGROPOULOS, S. I. et al. Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting, 2016.

WETTENGEL, J. Clean Energy Wire. **Clean Energy Wire**, 20 April 2017. Disponivel em: <<https://www.cleanenergywire.org/factsheets/very-brief-timeline-germanys-energiewende>>.

ZHONGYANG ZHAO, C. F. W. **Improvement to the Prediction of Fuel Cost Distributions Using ARIMA Model.** [S.l.]. 2018.