

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE CIÊNCIAS ECONÔMICAS
DEPARTAMENTO DE ECONOMIA E RELAÇÕES INTERNACIONAIS**

MATHEUS VIZZOTTO DOS SANTOS

**PROJEÇÃO DE DEMANDA ELÉTRICA COM
ALGORITMOS DE APRENDIZADO DE MÁQUINA**

**Porto Alegre
2022**

MATHEUS VIZZOTTO DOS SANTOS

**PROJEÇÃO DE DEMANDA ELÉTRICA COM
ALGORITMOS DE APRENDIZADO DE MÁQUINA**

Trabalho de conclusão submetido ao Curso de Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título Bacharel em Ciências Econômicas.

Orientador: Prof. Dr. Fernando Augusto Boeira Sabino da Silva

**Porto Alegre
2022**

CIP - Catalogação na Publicação

dos Santos, Matheus Vizzotto
Projeção de demanda elétrica com algoritmos de
aprendizado de máquina / Matheus Vizzotto dos Santos.
-- 2022.
60 f.
Orientador: Fernando Augusto Boeira Sabino da
Silva.

Trabalho de conclusão de curso (Graduação) --
Universidade Federal do Rio Grande do Sul, Faculdade
de Ciências Econômicas, Curso de Ciências Econômicas,
Porto Alegre, BR-RS, 2022.

1. Séries temporais. 2. Demanda. 3. Setor elétrico.
4. Algoritmos. 5. Aprendizado de máquina. I. da Silva,
Fernando Augusto Boeira Sabino, orient. II. Título.

MATHEUS VIZZOTTO DOS SANTOS

**PROJEÇÃO DE DEMANDA ELÉTRICA COM ALGORITMOS DE APRENDIZADO
DE MÁQUINA**

Trabalho de conclusão submetido ao Curso de Graduação em Ciências Econômicas da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título Bacharel em Economia.

Aprovada em: Porto Alegre, 05 de outubro de 2022.

BANCA EXAMINADORA:

Prof. Dr. Fernando Augusto Boeira Sabino da Silva – Orientador

UFRGS

Prof. Dr. Cristiano Lima Hackmann

UFRGS

Prof. Dr. Nelson Seixas dos Santos

UFRGS

*Sancta Dei Genitrici
in signum
gratitudinis et
filialis obedientia*

*O Maria, flos virginum,
velut rosa, vel lilium,
funde preces ad Filium
pro salute fidelium*

AGRADECIMENTOS

Agradeço à minha mãe, Vaine, por se sacrificar todos os dias pelo seu filho, e ao meu pai, José, que o fez enquanto pôde. Ao meu irmão, Marcelo, pelo carinho e apoio de sempre.

Ao meu orientador, Fernando, por confiar em mim na realização deste trabalho, e aos professores Cristiano e Nelson, por aceitarem avaliá-lo.

Ao meu gerente Adriano de Oliveira Martins, pelo apoio do início ao final desta jornada.

RESUMO

A indústria de energia elétrica requer uma grande rede de pessoas e estruturas conectadas trabalhando em harmonia para organizar os sistemas de geração, transmissão e distribuição que atendem os domicílios e negócios. Este trabalho tem como foco comparar modelos de séries temporais para prever a demanda de curto prazo no setor, cuja acuracidade permite gerência, agendamento e despacho de carga mais eficientes, estáveis e, consequentemente, com menores perdas operacionais para os provedores de energia. As abordagens com alisamento exponencial de Holt-Winters e com SARIMA foram comparadas aos algoritmos Prophet, TBATS, árvores de decisão, XGBoost, LightGBM e *ensemble learning* em um horizonte de quinze dias fora da amostra. As análises indicam uma superioridade destes em relação àquelas, em especial os modelos multivariado LightGBM, com erro percentual absoluto médio de 3,06%; Prophet, com 2,46%; e *ensemble learning*, com 1,60%.

Palavras-chave: Séries temporais. Setor elétrico. Projeção. Aprendizado de máquina. Algoritmos.

ABSTRACT

The electric power industry requires a massive network of people and structures working together in harmony to organize generation, transmission and distribution systems to provide for households and businesses. This paper aims to compare time series models on short-term electricity demand forecasting, whose accuracy leads to more efficient and stable managing, scheduling and dispatch of resources, thus resulting in fewer operational losses for the energy suppliers. Approaches such as the Holt-Winters exponential smoothing method and SARIMA were compared to the Prophet, TBATS, decision trees, XGBoost, LightGBM and ensemble learning algorithms on a fifteen-day out-of-sample forecast. The results indicated a superiority of these algorithms when compared to the first two models, specially the multivariate LightGBM, with a mean absolute percentage error of 3,06%; Prophet, with 2,46%; and ensemble learning, with 1,60%.

Keywords: Time series. Electricity market. Forecasting. Machine learning. Algorithms.

LISTA DE FIGURAS

Figura 1 - MAPE por mercado e por estação	21
Figura 2 - Série temporal estacionária	30
Figura 3 - Avaliação das funções FAC e FACP	31
Figura 4 - Visualização de uma árvore de decisão	34
Figura 5 - Variáveis analisadas	42
Figura 6 - FAC da série de carga elétrica	43
Figura 7 - APE dos melhores modelos	48
Figura 8 - Backtesting	50
Figura 9 - Distribuição dos erros em backtesting	51
Figura 10 - Sazonalidade anual da demanda elétrica na região sul do Brasil	59
Figura 11 - Sazonalidade semanal da demanda elétrica na região sul do Brasil	59
Figura 12 - Evolução da demanda elétrica na região sul do Brasil	60
Figura 13 - Projeção dos quatro melhores modelos	60

LISTA DE TABELAS

Tabela 1 - Transformação da série em um problema de regressão	44
Tabela 2 - Medidas de acurácia	45
Tabela 3 - Acurácia dos principais modelos	47
Tabela 4 - MAPE em $t+h$	49

Sumário

1 INTRODUÇÃO	11
1.1 OBJETIVO DESTE TRABALHO	14
1.2 ESTRUTURA DO TRABALHO	15
1.3 JUSTIFICATIVA DO ESTUDO	15
2 REVISÃO DE LITERATURA	16
3 METODOLOGIA	27
3.1 HOLT-WINTERS	27
3.2 SARIMA	28
3.3 PROPHET	32
3.4 ÁRVORES DE DECISÃO	33
3.4.1 Boosting	35
3.4.2 XGBoost	36
3.4.3 LightGBM	37
3.5 TBATS	38
3.6 ENSEMBLE LEARNING	39
4 DADOS, ESTRATÉGIAS DE PREVISÃO E ACURÁCIA	40
4.1 DADOS	40
4.2 ESTRATÉGIAS DE PREVISÃO	43
4.3 ACURÁCIA	44
5 RESULTADOS	47
5.1 BACKTESTING	49
6 CONSIDERAÇÕES FINAIS	52
REFERÊNCIAS	53
APÊNDICE A - ANÁLISE DA SÉRIE DE DEMANDA ELÉTRICA	59

1 INTRODUÇÃO

O processo de industrialização e o aumento na demanda por energia no Brasil ocorreram concomitantemente à maior participação do governo na produção de energia elétrica, marcada por um modelo de monopólio verticalizado. Na década de 1990, no entanto, frente à falta de competitividade, o sistema elétrico brasileiro passou por uma reestruturação com o projeto Re-SEB (Reestruturação do Setor Elétrico Brasileiro), que criou as bases para a atual composição do mercado deste setor. Com ele, também surgiu a Agência Nacional de Energia Elétrica (ANEEL), responsável pela regulamentação tarifária, pela contratação e pelo acesso aos sistemas de transmissão; o Operador Nacional do Sistema Elétrico (ONS), para coordenar as atividades do novo ambiente de negócios e com papel especial na operação do Sistema Integrado Nacional (SIN); a liberalização do mercado de energia elétrica; e a desverticalização das empresas desse setor. O racionamento de energia que ocorreu 2001 serviu de ensejo ao desenvolvimento do Novo Modelo do Setor Elétrico, que tinha como principais objetivos assegurar a expansão da oferta, a queda dos preços pela competição, o estímulo à eficiência econômica dos agentes e a inserção social (BRAGA, 2018). Para tanto, era fundamental a reestruturação do planejamento setorial e a formulação de um ambiente favorável ao investimento que possibilitasse a construção de novas usinas para garantir o abastecimento. Foi sancionada também a Lei 10.848 (2004), cujas regras para a comercialização de energia elétrica foram significativas para todos os agentes do mercado: as concessionárias, as permissionárias e as autorizadas de serviço público de distribuição de energia elétrica do SIN ficam obrigadas a garantir o atendimento à totalidade do mercado, mediante contratação regulada; é autorizada a criação da Câmara de Comercialização de Energia Elétrica (CCEE), para viabilizar a comercialização de energia elétrica, e a constituição do Comitê de Monitoramento do Setor Elétrico (CMSE), com a função no âmbito executivo de acompanhar e avaliar permanentemente a continuidade e a segurança do suprimento eletroenergético em todo o território nacional; e as concessionárias e autorizadas de geração são autorizadas a estabelecer contratos de compra e venda de energia elétrica pelo prazo de dez anos (BRASIL, 2004).

O aumento na livre concorrência e a desverticalização exigiram a reestruturação do sistema para consolidar uma indústria de energia elétrica de qualidade. A previsão de demanda tem um papel fundamental para o entendimento da oferta deste setor ao longo do tempo, para o planejamento e para a operação dos sistemas de energia elétrica. Como aponta Barros (2014), uma previsão acurada permite decidir com segurança pelo aumento de geração

de energia em um intervalo de tempo ou até mesmo a necessidade de expansão nas centrais elétricas dados o aumento da procura por energia e a expansão do sistema. Para atingir isso, trabalhos foram realizados com o objetivo de estimar o modelo com maior acurácia naquilo que Murto (1998) dividiu como projeções de curtíssimo, curto, médio e longo prazo. No curtíssimo prazo, incluem-se os modelos que consideram até minutos à frente; no curto prazo, de algumas horas até uma semana; para o médio prazo, semanas ou meses; para o longo prazo, desde alguns anos até décadas. Bordignon (2012) evidencia a importância das previsões de demanda de longo prazo na definição de estratégias de expansão na capacidade de atendimento aos consumidores, como construir novas plantas produtoras ou aumentar a capacidade das linhas de transmissão. A estimação correta no médio prazo auxilia na otimização de máquinas geradoras – e, consequentemente, em uma operação com custos minimizados –, na avaliação de instabilidade da rede elétrica e na negociação de contratos que reduzam os riscos financeiros (ROMERA *et al.*, 2008). Previsões de curto prazo, objeto deste trabalho, encontram sua utilidade nas decisões operacionais de geração, manutenção de carga e avaliação da segurança dos sistemas (BASHIR; EL-HAWARY, 2009), o que pode maximizar o uso de geradores térmicos para despacho econômico. Previsões de curtíssimo prazo, por fim, apoiam a tomada de decisão no gerenciamento da carga de demanda, na utilização imediata de novas unidades geradoras e na realocação de folgas de potência entre empresas.

A previsão acurada de demanda elétrica de curto prazo tem importância singular para que os provedores possam fazer frente ao desenvolvimento de mercados de energia com estratégias de precificação cada vez mais complexos, além de permitir gerência, agendamento e despacho de carga elétrica eficientes e estáveis (CHEN *et al.*, 2018). Para alcançar isso, vem sendo aplicada uma variedade de técnicas contidas basicamente em duas classes de modelos (RAFI *et al.*, 2021): os tradicionais e aqueles baseados em algoritmos de aprendizado de máquina (*machine learning*), que ganharam destaque em grande medida pela capacidade de capturarem a não-linearidade de séries de carga elétrica. Esta segunda categoria pode ser definida como um programa que

[...] aprende pela experiência E, com respeito a algum tipo de tarefa T e performance P, se sua performance P nas tarefas T, na forma medida por P, melhoram com a experiência E. (MITCHELL, 1997).

O aprendizado de máquina é um ramo da inteligência artificial relacionado ao desenvolvimento de algoritmos que podem estimar parâmetros e serem adaptados. Sua

principal diferença com a estatística tradicional é de que nesta geralmente se assume que há um modelo verdadeiro para o processo de geração dos dados e tenta-se estimar seus parâmetros, enquanto naquela se assume que os dados são gerados por um processo desconhecido, aproximado por diferentes algoritmos de aprendizado (FAWZY *et al*, 2021). Este grupo é ainda subdividido em algoritmos de aprendizado supervisionado, onde os dados têm a variável resposta identificada – como um problema de regressão – e aprendizado não supervisionado, onde os algoritmos trabalham com dados não identificados – como é o caso de *clusters* e análise de componentes principais (do inglês *Principal Component Analysis*, PCA).

Dentre os modelos tradicionais utilizados para previsão de curto prazo está o uso de regressão linear múltipla, de alisamento exponencial e de modelos autorregressivos integrados de médias móveis (ARIMA). Para os baseados em técnicas de inteligência artificial, nota-se o uso recorrente de lógica *fuzzy* (CAMPOS, 2008; ANDRADE, 2010; HAIHONG *et al*, 2022), de modelos híbridos (ROMERA, 2008; UNSIHUAY-VILA *et al*, 2010; VELASCO *et al*, 2018; RAFI *et al*, 2021; MASSAOUDI *et al*, 2021), de redes de funções de base radial (ROMERA, 2008; ALTRAN, 2010; RAFI *et al*, 2021) e de redes neurais artificiais (PAI; HONG, 2005; UNSIHUAY-VILA *et al*, 2010; MAEDA, 2021; VEERAMSETTY *et al*, 2022), que é a abordagem mais comum nos trabalhos para essa tarefa. Outros estudos aplicaram ainda modelos como regressão não paramétrica, autorregressão multiplicativa e processos autorregressivos de médias móveis com variável exógena (KWON *et al.*, 2020). Quanto aos fatores de maior influência na predição, Andrade (2010) aponta que as variáveis mais significativas são a carga, a temperatura e o preço da energia elétrica.

1.1 OBJETIVO DESTE TRABALHO

Buscou-se, ao longo deste trabalho, alcançar o modelo mais eficiente na projeção de demanda elétrica diária ao longo de 15 dias na região sul do Brasil, compreendido pelo sistema SIN como o sistema de atendimento aos estados do Rio Grande do Sul, Santa Catarina e Paraná, possibilitando um balanço entre as previsões de maior acurácia e os modelos com maior facilidade de interpretação. Inicialmente foi realizada uma análise exploratória dos dados, o que torna possível a identificação de componentes relevantes no pré-processamento dos dados e na escolha dos modelos. Com isso, passou-se à aplicação das técnicas preditivas utilizando os dados diários de 01/01/2000 a 16/05/2022 para projetar os

últimos quinze dias de maio de 2022, período mais recente disponível na data deste trabalho. Os modelos comparados foram:

- Holt-Winters
- SARIMA
- Árvores de decisão
- XGBoost
- LightGBM
- Prophet
- TBATS
- *Ensemble learning*

1.2 ESTRUTURA DO TRABALHO

O capítulo 2, de revisão da literatura, é composto pela exposição dos trabalhos já elaborados acerca da previsão de demanda elétrica, especialmente a de curto prazo, junto com os métodos aplicados e as medidas de acurácia obtidas. Tanto os estudos que realizaram uma análise univariada quanto aqueles que utilizaram mais de uma variável foram considerados.

Os modelos propostos neste trabalho constam no capítulo 3. No quarto capítulo serão realizadas uma análise da série temporal e das variáveis relevantes para o trabalho e uma exposição das estratégias de previsão em séries temporais e das medidas de acurácia utilizadas para comparação. No capítulo 5 serão apresentadas as projeções de cada metodologia da seção 3, que serão comparadas com a demanda real para um período de quinze dias. Por fim, no capítulo 6, mostram-se as principais conclusões no que diz respeito aos resultados que os algoritmos geraram.

O trabalho foi desenvolvido em código Python e pode ser acessado integralmente [neste repositório](#) do GitHub, que contém as seções de análise exploratória, modelagem e projeção da série temporal, além dos *scripts* necessários para a coleta dos dados. Nele também estão contidas as bibliotecas utilizadas para a estimação de parâmetros dos modelos apresentados na seção 3.

1.3 JUSTIFICATIVA DO ESTUDO

Uma projeção eficiente para a demanda elétrica é de fundamental importância para o setor. Pai (2005) ressalta como erros de previsão aumentam os custos de operação das empresas: uma superestimação faz com que se exija mais dos geradores desnecessariamente

para acumular reservas dispendiosas, enquanto uma demanda subestimada implica problemas de atendimento aos clientes. Bordignon (2012) aponta que a previsão de demanda de curto prazo auxilia na gerência da carga e leva à otimização do uso de potência, uma vez que tem impacto direto na partida e manutenção de unidades geradoras e na realocação de folgas do insumo entre empresas. A reduzida disponibilidade de estudos atualizados para esta finalidade na região sul do Brasil serve como motivação principal do trabalho, bem como a investigação de novas metodologias disponíveis para a análise.

2 REVISÃO DE LITERATURA

Como apontado na introdução deste trabalho, uma variedade considerável de métodos de previsão vem sendo utilizada na projeção de demanda elétrica para os horizontes de longo, médio, curto e curtíssimo prazos. No caso do Brasil, a Empresa de Pesquisa Energética (EPE), vinculada ao Ministério de Minas e Energia, desenvolve diversos estudos sobre a área. Dentre eles, a nota técnica “Metodologia: Modelo de Projeção de Demanda de Eletricidade” (2019) busca uma intermediação entre o que chamou de modelos econométricos e modelos analíticos, ou técnico-econômicos, para estimar o consumo de energia. A demanda aqui é tratada como função do crescimento da economia, da evolução da estrutura de renda nacional, do número de domicílios, de avanços tecnológicos, de condições climáticas, da arquitetura de grandes projetos industriais e de outros fatores estruturais. Fatores como a temperatura podem ter um efeito maior na série de curto prazo, enquanto aqueles como novos projetos industriais podem alterar o perfil de consumo de eletricidade no longo prazo – motivos pelos quais a metodologia de previsão de demanda de energia é costumeiramente alicerçada na formulação de premissas e cenários macroeconômicos. Além deste trabalho, também a nota “Metodologia: Projeção de Curva de Carga Horária” (2020) estuda a projeção de demanda elétrica ao discorrer sobre a utilização de modelos de redes neurais ANNSTLF (*Artificial Neural Network Short Term Load Forecast*) para análise de curto prazo horária. Neste último, como já é conhecido em outros setores em que modelos de *deep learning* são utilizados, as relações não-lineares podem ser projetadas mais eficazmente, mas isso vem acompanhado de uma perda significativa de capacidade de interpretação da metodologia; no planejamento de expansões de médio e longo prazos no sistema elétrico, esse é um elemento fundamental para que os tomadores de decisão possam avaliar diferentes cenários econômicos. Fatores conjunturais também podem ter um impacto na demanda não captado pelo modelo, como oscilações bruscas na temperatura e o custo da eletricidade, com papel ativo na definição do perfil da curva de carga. Além destes, alterações de hábitos de consumo e a adoção de novas tecnologias podem ter efeito sobre a carga horária em um prazo ainda mais longo.

Falco (2005) buscou nos modelos clássicos de Holt-Winters e Box & Jenkins uma alternativa ao método do ANNSTLF utilizado pelo ONS na projeção de demanda no subsistema Sudeste a partir da motivação de melhor inteligibilidade das projeções, uma vez que as redes neurais são consideradas uma “caixa preta”. A técnica de Holt-Winters foi utilizada para previsão de três formas: com ciclo diário e semanal; modelos considerando uma

constante de amortecimento para a tendência da série quando o horizonte de previsão fosse maior que um passo à frente; e modelos assumindo erros estruturados, já que a metodologia não abarca suficientemente séries de frequência horária. Foi observado que há um comportamento similar no perfil do consumo tanto de um determinado dia da semana quanto de uma determinada hora do dia, sendo necessário incorporar simultaneamente os ciclos diário e semanal na análise. Com a metodologia de Box & Jenkins, o melhor modelo obtido foi o SARIMA(0,1,0)(1,1,1)₂₄(2,1,1)₁₆₈, onde o sétimo e o décimo primeiro parâmetros, 24 e 168, se referem aos ciclos diários e semanais, respectivamente. O ANNSTLF, desenvolvido pelo *Electric Power Research Institute* (KHOTANZAD *et al*, 1998), fora utilizado pelo ONS principalmente para evitar investimentos desnecessários, repartir a capacidade de oferta entre as regiões do país, controlar a voltagem do sistema elétrico, coordenar hidrotérmicas e regular o fluxo das bacias. Esta arquitetura de redes neurais leva em conta também as variáveis de previsão horária de temperatura pelo próprio sistema e a temperatura horária real para melhor captura das oscilações de curto prazo com dados fornecidos por 16 aeroportos. Embora seja um algoritmo poderoso, a perda de interpretabilidade, o alto custo financeiro de utilização do software, a impossibilidade de adequar a arquitetura com autonomia e um maior erro percebido para as projeções da região sul mostravam motivos para que o programa fosse melhorado. Quantitativamente, foi utilizado o critério de Erro Absoluto Médio Percentual (do inglês *Mean Absolute Percentage Error*, MAPE) para comparar os modelos na avaliação dos dados do subsistema Sudeste. O modelo de Holt-Winters apresentou um MAPE de 0,34% para os dados utilizados no treinamento do modelo (*in-sample*) e de 1,60% nos dados fora da amostra (*out-of-sample*). A estimação feita via modelos ARIMA gerou um erro maior: 0,78% com os dados de treinamento (dentro da amostra) e 2,20% com os dados de teste (fora da amostra). O software ANNSTLF gerou o menor erro entre os três modelos, com 1,41% na partição fora de amostra. A perda de capacidade de generalização dos dois primeiros, quando testados sobre dados não utilizados no treinamento de algoritmos, é geralmente um indicativo de sobreajustamento, o que põe em dúvida a eficiência dos modelos estimados.

Uma abordagem diferente é trazida por Pai e Hong (2005) na projeção de carga elétrica regional em Taiwan. Partindo do pressuposto da existência de não-linearidade neste processo, os autores focam na utilização de *Recurrent Support Vector Machines* com algoritmos genéticos (RSVMG) para a projeção desta série em Taiwan. Os resultados indicaram para uma superioridade da nova implementação quando comparada aos modelos tradicionais de regressão e de redes neurais artificiais. Para a projeção de 1981 a 2000, o modelo RSVMG apresentou um MAPE médio entre as regiões de 1,42%; o de regressão, de 5,84%; e as redes

neurais artificiais, de 2,22%. A superioridade do primeiro foi evidenciada em relação tanto às medidas de erro quanto à aplicabilidade, já que um dos resultados do trabalho foi mostrar como essa metodologia proporciona uma convergência mais rápida do modelo.

Campos (2008) realizou uma comparação de diversos métodos para o consumo de longo prazo na cidade da Nova Inglaterra (EUA) e no estado de Minas Gerais (Brasil). Foram abordados os modelos ARIMA, NARIMA (ARIMA não-linear), Rede Neuro-Fuzzy (RNF) e Rede Neural Artificial (RNA). As projeções para Nova Inglaterra apresentaram MAPE de 0,204% de erro para a previsão de 12 passos à frente com o modelo RNA e de 0,129% para 60 passos com o ARIMA, enquanto a estimação do consumo no estado de Minas Gerais gerou erros da ordem de 0,075% para o modelo NARIMA na previsão de 12 meses à frente, de 0,361% para a RNA, de 0,431% para a RNF e de 1,956% para o modelo ARIMA. O estudo faz uma consideração importante para a utilização desta última metodologia no presente trabalho: com uma projeção de 125 meses à frente, o consumo convergiu para um valor fixo, ou seja, o modelo ARIMA é eficiente para previsões de curto prazo, enquanto sua utilização em horizontes de longo prazo converge para um valor fixo.

Romera (2008) propôs um novo modelo híbrido na projeção de médio prazo ao analisar os dados da Espanha de janeiro de 1975 a dezembro de 2002. A autora mostra que a imprecisão ao projetar variáveis climáticas e econômicas, comumente utilizadas como variáveis explicativas neste setor, é inevitável quando se considera um horizonte de tempo maior, o que leva, por sua vez, a mais erros nas projeções de demanda elétrica. Levando isso em consideração, são utilizadas como entradas do modelo apenas as defasagens da série histórica. Um dos pontos a que a autora chama atenção é o fato de que, embora a omissão de variáveis econômicas possa levar a inconsistências na identificação de tendência por não levar em conta aspectos como o desenvolvimento tecnológico, os fatores climáticos já estão de certa forma inseridos na série histórica, uma vez que o comportamento sazonal dessas variáveis são mais ou menos constantes para o mesmo mês em anos diferentes. A demanda mensal foi dividida em duas séries, uma de tendência e outra de flutuações. A primeira foi obtida ao aplicar um algoritmo de alisamento por médias móveis de doze defasagens; o segundo, subtraindo a tendência da série de demanda. Cada um desses conjuntos de dados foi dividido em uma partição de treino, de 276 observações, e outra de teste, com os últimos 60 meses, para medir a acurácia da projeção pelo erro percentual absoluto médio. Para projetar a série de flutuações foi utilizada uma análise espectral, enquanto redes neurais do tipo MLP e RBF (funções de base radial) serviram de algoritmo para previsão dos dados de tendência. Após isso, as duas séries foram agregadas, formando a previsão de demanda final. O resultado

foi um MAPE de 1,74% para a partição de teste, mostrando superioridade em relação a trabalhos que utilizaram apenas a metodologia ARIMA ou a de redes neurais para prever a tendência e as flutuações em uma única série.

Para obter um sistema de previsão de carga multinodal, ou seja, considerando vários pontos da rede elétrica, Altran (2010) empregou uma rede neural composta por módulos, onde cada um é constituído por redes do tipo MLP (*Multilayer Perceptron*, ou Perceptron Multicamadas) e estimado utilizando algoritmo de retropropagação. Como diferencial do trabalho, funções de base radial foram utilizadas como função de ativação no lugar da usual função sigmóide. Para todas as funções de ativação utilizadas, as projeções apresentaram resultado positivo com MAPE em torno de 2%.

O estudo de Andrade (2010) empregou o *Adaptive Neuro-fuzzy Inference System* (ANFIS) para prever a demanda de energia elétrica no curtíssimo prazo a fim de manter sua sintonia com a carga gerada pelo sistema. Trata-se de um sistema de inferência *fuzzy* implementado com redes neurais adaptativas que, por meio de um algoritmo de retropropagação, ajustam os parâmetros gerados por validação cruzada no algoritmo. O processo analisado foi a demanda medida em intervalos de cinco minutos ao longo de uma semana em subestações do estado de São Paulo: Cordeirópolis, Mogi-Guaçu, Ubatuba e Votuporanga. Alguns aspectos relevantes das séries temporais foram identificados, como baixo consumo nas primeiras horas da madrugada, um pico no entardecer e um comportamento intradiário distinto no final de semana, que identifica o consumo doméstico de energia. Os resultados de MAPE *in-sample* para cada uma das subestações tiveram média de 1,16%, representando, segundo o autor, uma melhora em relação ao que se encontrava na literatura até então.

Unsihuay-Vila *et al* (2010) propuseram um modelo híbrido denominado PREDICT2-ES, que se baseia em dinâmicas não-lineares com uma etapa de treinamento onde cinco parâmetros são otimizados pelo algoritmo de computação evolucionária, tendo como uma das contribuições a capacidade de capturar padrões de demanda e preço nas séries temporais de New England (Estados Unidos), Alberta (Canadá) e Espanha. Como ressaltam os autores, a previsão acurada do preço *spot*¹ é de grande importância para a operação do sistema elétrico, especialmente em mercados desregulados, já que os produtores podem aplicar estratégias de maximização de lucros de minimização de riscos enquanto os consumidores podem alocar compras entre contratos bilaterais de longo prazo e mercados de

¹ Preço do mercado de curto prazo de energia elétrica, onde se negociam contratos com períodos e montantes preestabelecidos.

curto prazo. Embora a literatura na previsão de preços do setor elétrico tenha semelhanças com os métodos aplicados sobre a demanda – ou seja, regressão múltipla, ARIMA, redes neurais e lógica *fuzzy* –, aqui a metodologia GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*) também ganha espaço. Para verificar a eficiência da abordagem proposta, os resultados, segmentados no âmbito de demanda e preço para cada uma das regiões e também no de estações do ano, são comparados com aqueles dos modelos ARIMA e redes neurais artificiais (ANN). Como mostra a figura 1, o modelo proposto apresentou um desempenho superior aos outros dois, embora os erros de previsão baseada nas estações do ano tenham apresentado maior variância frente àquela segmentada por mercado.

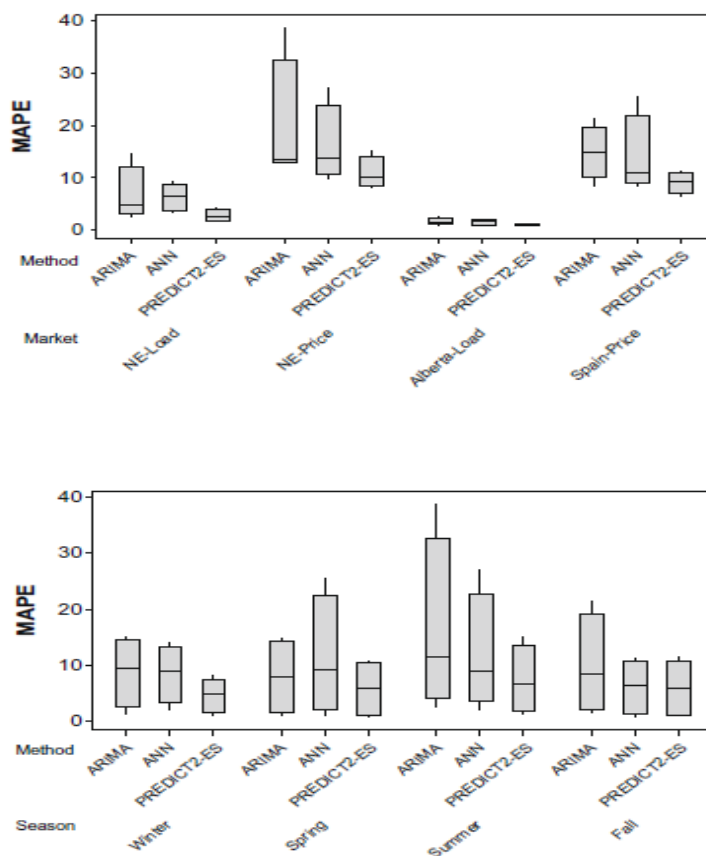


Figura 1: MAPE por mercado e por estação do ano. Extraído de Unsihuay-Vila *et al* (2010).

Um modelo multivariado foi proposto por Bordignon (2012) para prever a carga elétrica de curtíssimo prazo de uma Pequena Central Hidrelétrica (PCH) no Rio Grande do Sul com base na influência de variáveis climáticas. A criação de um modelo probabilístico discreto a partir da classificação dos dados históricos em um mapa auto-organizável alcançou a medida de 3,46% de MAPE para o horizonte de um dia útil do mês de dezembro de 2011.

Barros (2014) fez uma comparação entre a metodologia Box & Jenkins, Redes Neurais e Máquinas de Suporte Vetorial (SVM) na previsão de carga de curto prazo. Quando comparado aos dados reais, a estimação de um modelo SARIMA (2,0,0)(1,0,0)₂₄ gerou um MAPE de 3,01%, enquanto as Redes Neurais chegaram a 1,53% e as SVM em 2,09%. A normalização dos dados utilizados como entradas dos modelos e a consideração de feriados na série temporal foram aspectos relevantes no trabalho para a minimização dos erros de previsão.

Uma metodologia baseada em *ensemble learning*² foi aplicada por Fan *et al* (2014) a partir da técnica de mineração de dados para prever o consumo de energia um dia à frente e os picos de demanda energética em construções de Hong Kong. Considerando uma série com intervalo de 15 minutos ao longo de um ano, inicialmente é realizada uma análise de valores anormais (*outliers*) pelos métodos de *clusters* e *generalized extreme studentized deviate* (GESD) a fim de remover observações prejudiciais à estimação dos modelos para, então, por meio do método de eliminação recursiva de atributos (RFE), selecionar as variáveis explicativas mais relevantes para oito algoritmos preditivos diferentes. Essa técnica funciona pela estimação de um modelo com todas as variáveis candidatas, as quais são recursivamente classificadas de acordo com um critério que avalia a importância da variável na predição e elimina as menos significativas. A razão de se utilizar um *ensemble* advém do fato de modelos individuais terem suas vantagens e desvantagens, que são equilibradas pela combinação ótima de pesos atribuídos aos algoritmos. Para gerar essa combinação no *ensemble*, um algoritmo genético (GA) foi utilizado nos modelos de regressão linear múltipla (MLR), ARIMA, *support vector regression* (SVR), *random forests* (RF), perceptron multicamada (MLP), *boosting tree* (BT), *multivariate adaptive regression splines* (MARS) e *k-nearest neighbors* (KNN), sendo os dois primeiros considerados individualmente como referências de performance. Os resultados foram um MAPE de 2,32% para a previsão um dia à frente e de 2,85% para o pico de demanda.

A previsão de consumo elétrico no médio-longo prazo foi realizada em outro trabalho (OLIVEIRA, E.; OLIVEIRA; F., 2018) com uma metodologia inovadora para os dados mensais em diferentes países da Organização para a Cooperação e Desenvolvimento Econômico (OCDE). É realizada uma combinação de diferentes técnicas de decomposição e *bootstrap aggregation* para simular novas séries e, depois, projetá-las com uma variedade de métodos. Os resultados são por fim combinados em uma única projeção por média simples e mediana e comparados com os dados reais de partição de teste dos dados. O método utilizado

² Combinação de modelos preditivos a fim de obter melhor precisão.

para obter a tendência, a sazonalidade e os componentes restantes da série de demanda foi o STL (*Seasonal-Trend decomposition using Loess*), uma sequência de operações de alisamento que utilizam regressões ponderadas localmente; dentre as suas vantagens estão a possibilidade de lidar com qualquer tipo de sazonalidade e com *outliers* ao estimar os componentes de tendência e ciclo. Os métodos empregados nas novas séries geradas pelo *bootstrap aggregation* foram os de alisamento exponencial, como os modelos aditivo e multiplicativo de Holt-Winters, e SARIMA, cujas estimações representaram um aumento na acurácia da projeção de demanda elétrica.

Outro modelo híbrido foi utilizado por Velasco *et al* (2018) na previsão de carga elétrica de curto prazo em uma instalação de Mindanao, nas Filipinas, que até então se valia de um modelo de regressão linear com MAPE superior a 5%. A combinação de um modelo ARIMA com uma rede neural artificial se mostrou promissora para prever a demanda um dia à frente, gerando um MAPE de 4,09%, enquanto os modelos considerados isoladamente apresentaram erros da ordem de 5% e 17% respectivamente. Os dados foram coletados com intervalos de 15 minutos e continham interrupções agendadas e não planejadas, resultando em observações sem valor de consumo. Para contornar possíveis impactos negativos nas estimações, fez-se uma transformação na série de acordo com o que pesquisas do setor usualmente recomendam: utilizar o valor do mesmo momento no dia anterior para preencher valores vazios. O modelo ARIMA (8,1,2) é então estimado e as suas projeções armazenadas como a parte linear do resultado. Os seus resíduos são utilizados pela rede de perceptron multicamadas com função de ativação tangente hiperbólica, cuja arquitetura consiste em uma camada de entrada com 24 neurônios – correspondente ao horizonte de projeção, ou seja, 24 horas –, uma camada oculta com 17 neurônios e uma camada de saída com 1 neurônio. As projeções da rede neural artificial, parte não linear dos resultados, e do modelo ARIMA são agregadas por soma e avaliadas pelas medidas MAPE e MSE (Erro Quadrático Médio) na partição de teste dos dados, que abrange o dia 21 de outubro de 2014. Também foram realizadas projeções de 2, 3 e 4 dias à frente para verificar se o mesmo modelo mantinha sua acurácia, mas os erros aumentaram junto com o horizonte de predição, chegando a 8,60%. Com isso, os autores sugerem que o modelo seja treinado novamente toda vez que for utilizado para prever o consumo de carga elétrica no dia seguinte.

Zheng *et al* (2019) desenvolveram um modelo para projeção de carga residencial de curto prazo. A metodologia avaliada foi a implementação de um filtro de Kalman, desenvolvido inicialmente em 1960 com base na teoria de probabilidade bayesiana e em cadeias de Markov, em dados do Reino Unido e Estados Unidos. Seu diferencial é uma

abordagem *bottom-up*, ou seja, com a agregação de projeções divididas em subsistemas, como dispendio em ar condicionado e iluminação na residência. As projeções foram comparadas pela métrica de erro percentual absoluto médio simétrico (sMAPE) com os resultados de outros dois modelos: *naive* e rede LSTM. O primeiro se refere a um modelo simples geralmente utilizado como referência nos trabalhos de séries temporais, pois considera o valor da última observação na série temporal como os valores futuros estimados. Os autores apontam como pontos sensíveis da metodologia adotada a acurácia, uma vez que o comportamento dos residentes, que varia de pessoa para pessoa, tem um impacto direto na variabilidade da carga elétrica no local. No entanto, o resultado foi um valor de sMAPE 49% inferior àquele do modelo LSTM.

Em um estudo do consumo de curto prazo de energia elétrica de uma microrrede de prédios em um polo tecnológico brasileiro, Maeda (2021) desenvolveu um modelo de redes neurais artificiais ANFIS e transformada *wavelets*. Na comparação das diversas famílias de *wavelets* utilizadas na projeção – como as de Meyer, Haar, Daubechies, Symlets e Coiflets – os melhores modelos alcançaram um erro percentual médio absoluto de 6,7% no período de uma semana.

Rafi *et al* (2021) propuseram uma abordagem inovadora com um modelo híbrido de rede neural convolucional (CNN) e *Long Short-Term Memory* (LSTM) aplicado à série semi-horária de carga elétrica em Bangladesh para o período de janeiro de 2014 a dezembro de 2019, comparando-o com o desempenho de outras abordagens já existentes na literatura: uma rede LSTM, uma rede de função de base radial (RBF) e um modelo XGBoost (*Extreme Gradient Boosting*). Os dados foram separados em um conjunto de treino, que vai de janeiro de 2014 a dezembro de 2018, e em outro de teste, que contempla o ano de 2019. Avaliando o MAPE nesta partição, a combinação entre CNN e LSTM mostrou uma projeção mais acurada que a rede LSTM isoladamente, embora ambos os modelos tenham sido capazes de captar os componentes da série temporal. Para os horizontes mensal, semanal e de 48 e 24 horas, o erro médio foi de 7,99% para o modelo LSTM; de 4,13% para o modelo híbrido proposto pelos autores; de 9,40% para as redes de função de base radial; de 11,56% para o XGBoost.

A metodologia de *ensemble* também foi explorada por Guo *et al* (2021) ao implementar um algoritmo de fusão entre três métodos de aprendizagem de máquina para previsão de carga elétrica de curto prazo: *support vector machine* (SVM), *random forest* (RF) e rede neural *Long Short-Term Memory* (LSTM). Os autores chamam atenção para as vantagens e desvantagens para cada um desses modelos, que é a motivação para o uso de um algoritmo composto, como a capacidade da técnica de SVM evitar problemas de ótimo local na

estimação dos coeficientes e de funcionar com pequenas amostras de dados, assim como boa capacidade de generalização, ao passo que tem como pontos negativos o custo computacional excessivo e a insensibilidade a dados faltantes em amostras maiores. O *random forest* consegue lidar com empecilhos de sobreajustamento em um tempo reduzido, mas é pouco interpretável. A rede LSTM, por fim, lida com problemas de ótimos locais e de sobreajustamento, mas é dispendioso do ponto de vista computacional. Um passo de relevância no estudo foi o tratamento de dados, mais especificamente das observações anormais, que podem ser geradas pelos medidores com falha técnica ou por uma queda repentina de carga – geralmente em decorrência de uma grande variação na demanda por parte da indústria. Os dados utilizados consistiam de 5.760 observações com intervalo de 15 minutos, sendo as últimas 96 utilizadas na partição de teste da série. A arquitetura da rede LSTM foi desenvolvida com três camadas ocultas de 10, 15 e 5 neurônios, respectivamente, e uma camada de saída com um único neurônio. O resultado foi um MAPE de 23,4% para o algoritmo SVM, de 3,6% para o *random forest*, de 5,3 para a rede LSTM e 2,8% para o modelo de *ensemble* proposto.

Um modelo híbrido empilhado que combina o algoritmo Light Gradient Boosting Machine (LGBM), Extreme Gradient Boosting Machine (XGB) e rede de perceptron multicamada (MLP) foi desenvolvido por Massaoudi *et al* (2021). O método realiza a geração de metadados pelos dois primeiros modelos para, com o último, gerar a previsão final. Proposto pela Microsoft (KE *et al.*, 2017), o LGBM aprimora as capacidades do modelo de Gradient Boosted Decision Trees (GBDT) no que se refere à eficiência computacional e ao consumo de memória enquanto preserva a acurácia elevada e previne o sobreajustamento. Também foi realizada uma análise de otimização de nove hiperparâmetros com cinco técnicas diferentes, mais especificamente o número de estimadores, a taxa de aprendizado, a profundidade máxima da árvore nos modelos LGBM e XGB, o número máximo de iterações e o tamanho da camada oculta da rede MLP. As métricas de acurácia utilizadas foram o RMSE, MAE, MAPE, R^2 e MSLE (Mean Squared Logarithmic Error) sobre 20% de dois conjuntos de dados horários em localizações diferentes. O primeiro em Johor, Malásia, entre 2009 e 2011, e o segundo em New England, de 2003 a 2014. Analisados individualmente, o modelo LGBM teve um R^2 de 89,1%, a rede MLP de 87,9% e o XGB de 86,5%. Para o modelo empilhado, esse valor foi de 94,3% com um MAPE de 2,7%.

Haihong *et al* (2022) aplicaram um modelo híbrido de *deep learning* considerando o efeito acumulado da temperatura na carga elétrica de uma região no leste da China. Considerando os fatores que comumente influenciam a carga de curto prazo e analisando a

correlação, é estabelecido o modelo com variáveis de temperatura e o seu efeito acumulado. Depois, a rede convolucional temporal (TCN) extrai e constrói o vetor de variáveis explicativas da série temporal, cujos resultados são utilizados como novas entradas da Deep Neural Network (DNN), passo final da projeção. Os dados de carga contemplam intervalos de 15 minutos e o objetivo foi utilizar os dados dos últimos 30 dias para prever a carga ao longo de um dia à frente. Para ajustar outliers, foi utilizado o método de alisamento horizontal; os valores faltantes foram preenchidos com observações adjacentes. A consideração do efeito cumulativo advém do fato de que, sob cenários de temperatura alta e contínua, a carga terá um efeito anormal de crescimento, ou seja, haverá um índice de carga significativamente diferente sob o mesmo valor de temperatura. O trabalho utiliza inferência fuzzy Mamdani para fazer essa correção, que passa a representar o efeito de temperaturas contínuas sobre a variável explicada. A aplicação do híbrido TCN-DNN visa a contornar os problemas observados em modelos RNN e LSTM, como o alto consumo de memória dos computadores e a baixa capacidade de lembrar de informações históricas. O modelo TCN consegue extrair os padrões mais complexos dos dados e obter os fatores mais importantes da série, mas não captura suficientemente informações de dados não-sequenciais, motivo pelo qual se utiliza a capacidade da DNN de processar dados não cronológicos como complemento à projeção. Os modelos TCN e DNN considerados individualmente tiveram melhor performance que o LSTM e o modelo com correção de temperatura apresentou uma superioridade de 0,91% em relação ao modelo TCN-DNN sem esse processamento, resultando em uma acurácia final de 97,92% e maior eficiência na otimização dos parâmetros.

Para prever a carga de curto prazo em uma subestação elétrica na Índia por meio de uma metodologia com custos computacionais reduzidos, Veeramsetty *et al* (2022) utilizaram uma análise de componentes principais (PCA) em conjunto com uma rede neural recorrente (RNN). A abordagem considera o impacto temporal das últimas três horas, assim como dos últimos três dias e três semanas no mesmo horário como variáveis de entrada da rede neural. No entanto, a PCA, que comprime o conjunto de dados original em um subespaço com o objetivo de preservar a maior variância explicada mantendo o mínimo de variáveis, extrai as características mais importantes destas nova defasagens e as transforma em seis, responsáveis por aproximadamente 90% da variância total. Essa redução no número de variáveis corresponde a uma menor camada de entrada da rede neural e um menor tempo de processamento ao mesmo tempo em que mantém quase a mesma acurácia dos dados originais. Quando comparada a outros modelos utilizados na projeção de carga de curto prazo, como redes neurais artificiais (ANN), modelos de regressão e redes LSTM, a rede neural recorrente

apresentou boa acurácia, além de ter melhores resultados que modelos lineares AR, MA, ARMA e ARIMA. O modelo cujo pré-processamento utilizou a PCA teve praticamente a mesma acurácia que o modelo com as nove variáveis originais, mas com menor complexidade.

Velasquez *et al* (2022) fizeram uma análise da projeção de demanda elétrica mensal no Brasil e nas quatro regiões classificadas pelo ONS, ou seja, Norte, Nordeste, Sudeste e Sul com os modelos de regressão com sazonalidade, alisamento exponencial de Holt-Winters e ARIMA. Os dados abrangem a carga de energia do período de janeiro de 2002 a dezembro de 2020. Embora o ONS disponibilize dados desde janeiro de 1999, os autores optaram por excluir o período da crise de blackout ocorrida em 2001, quando cortes forçados na energia elétrica foram determinados pelo governo brasileiro em algumas regiões e também iniciou a gestão de programas voltados à redução do consumo de eletricidade. Para comparar os três modelos, foram utilizadas as métricas de erro médio, erro médio absoluto, erro quadrático médio, erro percentual médio e erro percentual absoluto médio. A regressão com sazonalidade se mostrou superior aos outros modelos com o menor erro percentual em todos os cenários, embora tenha apresentado, à exceção da região Sul, valores superiores a 10%, chegando a 34,8% para a região Norte. A combinação média dos três modelos gerou melhores resultados, o melhor sendo de 0,99% para a região Norte.

3 METODOLOGIA

Esta seção apresenta os modelos univariados, que utilizam uma única série temporal como preditora, e multivariados, com duas ou mais, aplicados no trabalho para projetar a série de demanda elétrica de curto prazo na região sul do Brasil. Há um foco em modelos capazes de capturar componentes como sazonalidade e tendência, relevantes nos dados utilizados (FALCO, 2005). Também são apresentados conceitos resgatados no capítulo 4 para identificação de aspectos relevantes à modelagem da série temporal, como funções de autocorrelação e testes de estacionariedade.

3.1 HOLT-WINTERS

O modelo de Holt-Winters, baseado em alisamento exponencial, é uma extensão do modelo de Holt aplicada a séries com tendência ou com tendência e sazonalidade (ATHANASOPOULOS; HYNDMAN, 2021). O método pode ser descrito como

$$\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)}, \quad (1.a)$$

$$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_t + b_{t-1}), \quad (1.b)$$

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}, \quad (1.c)$$

$$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}, \quad (1.d)$$

sendo ℓ_t uma equação do nível da série; b_t a de tendência; s_t a do componente de sazonalidade; $\hat{y}_{t+h|t}$ a de projeção; α , β^* , e γ os coeficientes de alisamento de cada uma, respectivamente, e m é a frequência da série, ou seja, o período de sazonalidade (dados mensais têm $m=12$, por exemplo). Nessa equação foi considerada a versão aditiva do modelo, ou seja, quando as variações sazonais são relativamente constantes ao longo do tempo. Quando esse não é o caso, utiliza-se então o modelo multiplicativo para capturar variações sazonais proporcionais ao nível da série:

$$\hat{y}_{t+h|t} = (\ell_t + hb_t)s_{t+h-m(k+1)}, \quad (2.a)$$

$$\ell_t = \alpha \frac{y_t}{s_{t-m}} + (1 - \alpha)(\ell_{t-1} + b_{t-1}), \quad (2.b)$$

$$b_t = \beta^* (\ell_t - \ell_{t-1}) + (1 - \beta^*) b_{t-1}, \quad (2.c)$$

$$s_t = \gamma \frac{y_t}{\ell_{t-1} + b_{t-1}} + (1 - \gamma) s_{t-m}, \quad (2.d)$$

3.2 SARIMA

Com origem no trabalho de Box e Jenkins (1970), a metodologia busca descrever as autocorrelações de uma série temporal, lidando, em acréscimo ao modelo ARIMA não sazonal, com o componente de sazonalidade. O termo “AR” se refere a um modelo autorregressivo, ou seja, uma regressão de uma variável sobre ela mesma, que apresenta grande flexibilidade para lidar com diferentes padrões de séries temporais (ATHANASOPOULOS; HYNDMAN, 2021) e pode ser representado pelo processo estocástico AR(p)

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t = \sum_{j=1}^p \phi_j y_{t-j} + \varepsilon_t, \quad (3)$$

em que c é uma constante, ε_t é um ruído branco, ou seja, uma série com média zero, variância constante e sem correlação com outras observações da sequência, e y_{t-i} é uma defasagem da variável y_t .

O termo “MA” se refere a um modelo de médias móveis, onde o valor de cada y_t é como uma média móvel ponderada dos erros de previsão passados representada pela equação 4.

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} = \mu + \sum_{j=0}^q \theta_j \varepsilon_{t-j}. \quad (4)$$

O processo autorregressivo de médias móveis é formado pela combinação desses dois modelos, sendo um ARMA(p, q) definido como

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=0}^q \theta_j \varepsilon_{t-j}. \quad (5)$$

A estacionariedade fraca (de segunda ordem) é necessária para que a série temporal tenha uma representação dentro desta classe de modelos como demonstrado pelo teorema de Wold (SHUMWAY; STOFFER, 2016). A série y_t é fracamente estacionária se:

1. $E |y_t|^2 < \infty$
2. $E(y_t) = \mu$, para todo $t \in \mathbb{Z}$
3. $E(y_t - \mu)(y_{t-j} - \mu) = \gamma_j$

A primeira condição se refere ao segundo momento não centrado, que deve ser finito. A segunda postula que a média não depende do tempo t , ainda que a distribuição da variável aleatória seja alterada ao longo do tempo. A terceira, que a autocovariância não depende do tempo, mas apenas da distância temporal entre as observações.

A diferenciação é a subtração de observações consecutivas em uma série temporal e representa a ordem do termo “I” (*Integrated*). No caso de um modelo com tendência

$$x_t = \mu_t + y_t, \quad (6)$$

onde x_t são as observações, y_t é uma série estacionária e μ_t representa o componente estocástico de tendência com *drift* δ e ruído branco w_t

$$\mu_t = \delta + \mu_{t-1} + w_t, \quad (7)$$

diferenciar a série torna o processo estacionário, como pode ser visto na equação 8.

$$x_t - x_{t-1} = (\mu_t + y_t) - (\mu_{t-1} + y_{t-1}), \quad (8)$$

$$x_t - x_{t-1} = \delta + w_t + y_t - y_{t-1}$$

Se a estacionariedade não for verificável após a primeira diferenciação, pode ser realizada uma segunda, onde $\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$. No primeiro caso, caracteriza-se a série como I(1); no segundo, como I(2). Índícios de uma série estacionária podem ser verificados na figura 2, onde as observações giram em torno da média zero e não apresentam alteração na variância.

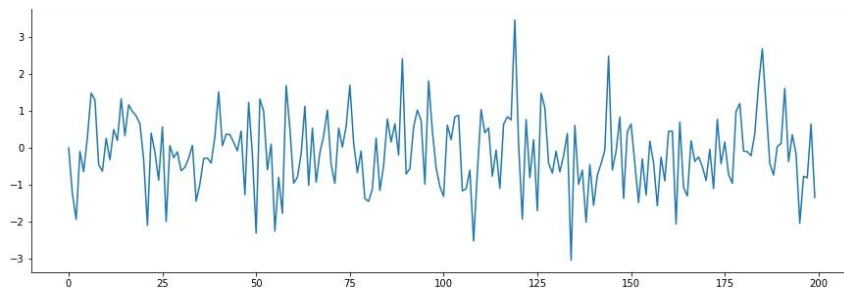


Figura 2: Série temporal estacionária.

No entanto, a simples inspeção visual de um processo é insuficiente para diferenciá-lo entre estacionário ou não estacionário. Testes estatísticos de raiz unitária, como os de Dickey-Fuller, Dickey-Fuller Aumentado (DICKEY; FULLER, 1979; DICKEY; FULLER,

1981), Phillips-Perron (PHILLIPS; PERRON, 1988) e KPSS (KWIATKOWSKI *et al*, 1992) foram desenvolvidos para fazê-lo com maior assertividade.

Para identificar a ordem q de um processo de médias móveis, utiliza-se a função de autocorrelação (FAC). No caso de um $MA(q)$, a autocorrelação é zero (trunca) a partir da defasagem $q+1$, enquanto para processos $AR(1)$, $AR(p)$ e $ARMA(1,1)$, por exemplo, há um decaimento progressivo. Para a identificação da ordem p , por outro lado, mantém-se apenas a correlação direta pura entre duas observações pela função de autocorrelação parcial (FACP). Para avaliar a autocorrelação de uma observação com suas defasagens, são empregados gráficos como os da figura 3, que representa um processo $MA(2)$. Um modelo $ARMA$ apresentará decaimento exponencial tanto no gráfico de FAC quanto no de FACP, dificultando a identificação. A regra usual para esses casos é a de selecionar o menor critério de informação de todas as combinações de p e q para o processo, tendo como cenário limite o modelo em que os coeficientes de autocorrelação deixam de ser estatisticamente significativos.

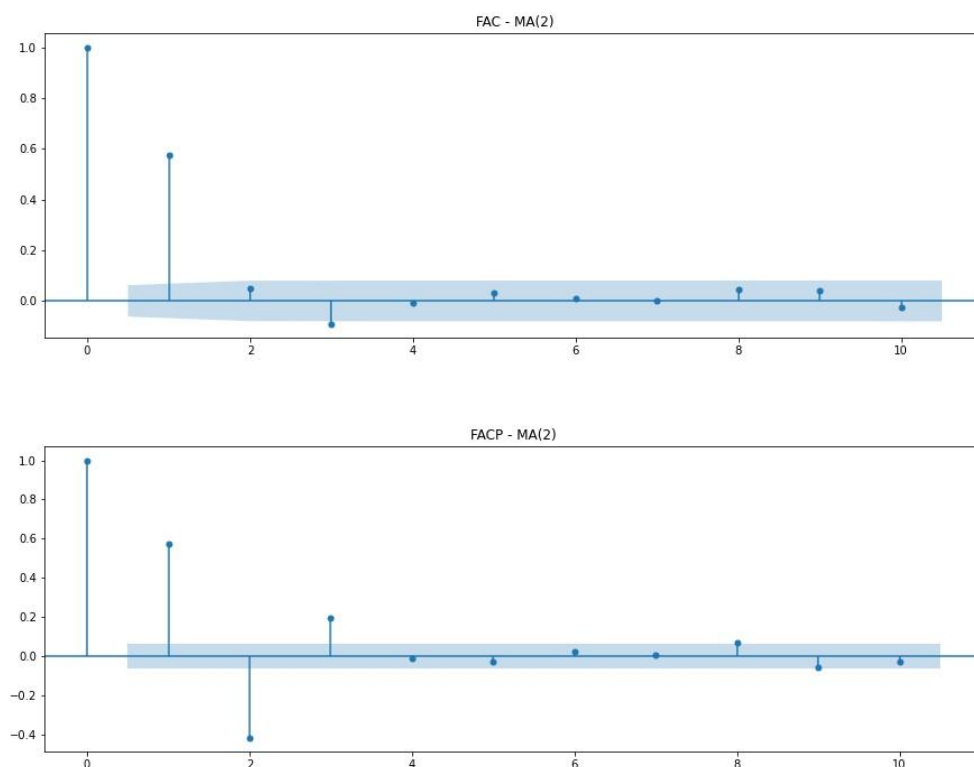


Figura 3: Avaliação das funções FAC e FACP.

A abordagem desenvolvida até aqui possibilita a estimação de diversos padrões em séries sem sazonalidade. Para contemplar este componente, o modelo SARIMA inclui

componentes autorregressivos e de médias móveis sazonais representados por ARIMA (p, d, q) x (P, D, Q)_m: “m” representa a frequência do processo (anual, mensal, diário etc.), e P, D e Q se referem à ordem dos processos autorregressivo, de diferenciação e de médias móveis respectivamente. A equação do modelo pode ser descrita como

$$\Phi_p(B^S)\phi(B)\nabla_S^D\nabla^d x_t = \delta + \Theta_Q(B^S)\theta(B)w_t, \quad (9)$$

em que $\phi(B)$ e $\theta(B)$ são respectivamente os componentes autorregressivo e de médias móveis usuais; $\Phi_p(B^S)$ e $\Theta_Q(B^S)$, os componentes autorregressivos e de médias móveis sazonais de ordem P e Q; $\nabla^d = (1 - B)^d$; $\nabla_S^D = (1 - B^S)^D$; e B é o operador de defasagem $Bx_t = x_{t-1}$. A identificação dos parâmetros P, D e Q segue o padrão da parte não sazonal, com a diferença de serem identificados nas defasagens sazonais dos gráficos de FAC e FACP.

Box e Jenkins (1970) propuseram uma abordagem de três estágios para selecionar um modelo apropriado à estimação e projeção de séries temporais, que consiste na identificação, na estimação e no diagnóstico (ENDERS, 2014). A primeira etapa consiste na análise visual do processo, assim como da função de autocorrelação e da função de autocorrelação parcial; na segunda os coeficientes são estimados e o modelo com melhor ajuste, comumente comparado aos outros pelo critério de informação de Akaike (AIC) e pelo critério de informação bayesiano (BIC), é selecionado; na etapa de diagnóstico, por fim, é verificado se os resíduos do modelo estimado se caracterizam como ruído branco. Para este estágio da análise, em especial, os testes de Box-Pierce (1970) e de Ljung-Box (1978) são largamente utilizados para verificar a existência de autocorrelação nos resíduos de modelos ARIMA estimados.

3.3 PROPHET

Apresentado por Taylor e Letham (2018), o algoritmo que ficou conhecido como Prophet foi desenvolvido para fazer projeções de séries temporais univariadas pela implementação de um modelo aditivo capaz de capturar componentes como tendência, sazonalidade e feriados. Desenvolvido tendo em vista o ambiente de negócios, pode contemplar séries horárias, diárias, semanais e mensais com múltiplas sazonalidades, efeitos de feriados que ocorrem em intervalos irregulares mas conhecidos com antecedência, valores faltantes, *outliers*, quebras estruturais e tendências não lineares restritas por um limite natural.

Uma das vantagens apontadas pelos autores é a superioridade do algoritmo em relação à aplicação de um modelo SARIMA, que, embora capture componentes sazonais não identificados pelo ARIMA convencional, apresenta um custo computacional significativamente superior para sua estimação.

O algoritmo é composto por quatro componentes, entre os quais está a estimação em partes da tendência da curva de crescimento linear ou logística, obtida pela seleção de pontos de mudança na série temporal; o componente de sazonalidade anual estimado por séries de Fourier; o de sazonalidade semanal, por variáveis *dummy*; e uma lista de feriados relevantes. O modelo de série temporal é composto na equação 10, onde $g(t)$ é uma função de tendência que captura mudanças não periódicas, $s(t)$ representa mudanças periódicas e $h(t)$ os efeitos de feriados e ϵ_t mudanças não identificadas pelo modelo com distribuição normal.

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (10)$$

O modelo de tendência linear com pontos de mudanças é descrito na equação 11, no qual k representa a taxa de crescimento, $a(t) \in \{0,1\}^S$ é um vetor contendo a taxa base adicionada a todos os ajustamentos até o ponto S, δ é a taxa de ajustamento, m é o parâmetro de deslocamento e γ corresponde a $-s_j \delta_j$.

$$g(t) = (k + a(t)^T \delta)t + (m + a(t)^T \gamma) \quad (11)$$

O componente de sazonalidade, como já comentado, vale-se de séries de Fourier para identificar múltiplos padrões na série temporal, como na equação 12, onde P é o período regular esperado da série.

$$s(t) = \sum_{n=1}^N \left(\cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right) \quad (12)$$

Por fim, é estabelecida uma função indicadora para representar se o tempo t ocorre no feriado i e atribui a cada um o parâmetro k_i , que é a variação correspondente na série, por meio de uma matriz de regressores $Z(t) = [1(t \in D_1), \dots, 1(t \in D_L)]$ em que

$$h(t) = Z(t)k. \quad (13)$$

3.4 ÁRVORES DE DECISÃO

Modelos baseados em árvores de decisão estão entre aqueles mais utilizados e de maior acurácia nos métodos de aprendizado de máquina supervisionado e são superiores aos modelos tradicionais na identificação de relações não lineares, além de se adaptarem com facilidade a problemas de classificação e de regressão (FAWZY *et al.*, 2021). Valendo-se disso, Navin (2015), por exemplo, utilizou árvores de decisão para desenvolver um modelo de previsão dos preços do ouro baseado em seu histórico e comparou os resultados com o modelo SVM, concluindo que o tempo de processamento e o erro quadrático médio do primeiro são menores que os do segundo.

Árvores de decisão dividem o espaço das variáveis explicativas em um conjunto de retângulos para depois estimar um modelo simples em cada um, como uma constante (HASTIE *et al.*, 2009). Um exemplo de uma árvore de decisão utilizada para estimar uma série temporal com duas defasagens pode ser visualizada na figura 4: do *nó raiz*, que divide as observações entre aquelas com valor menor ou igual a 6,875 em $t-1$ à esquerda e aquelas com valor maior à direita, até os *nós folhas*, na parte inferior do gráfico, que não se dividem em mais subgrupos.

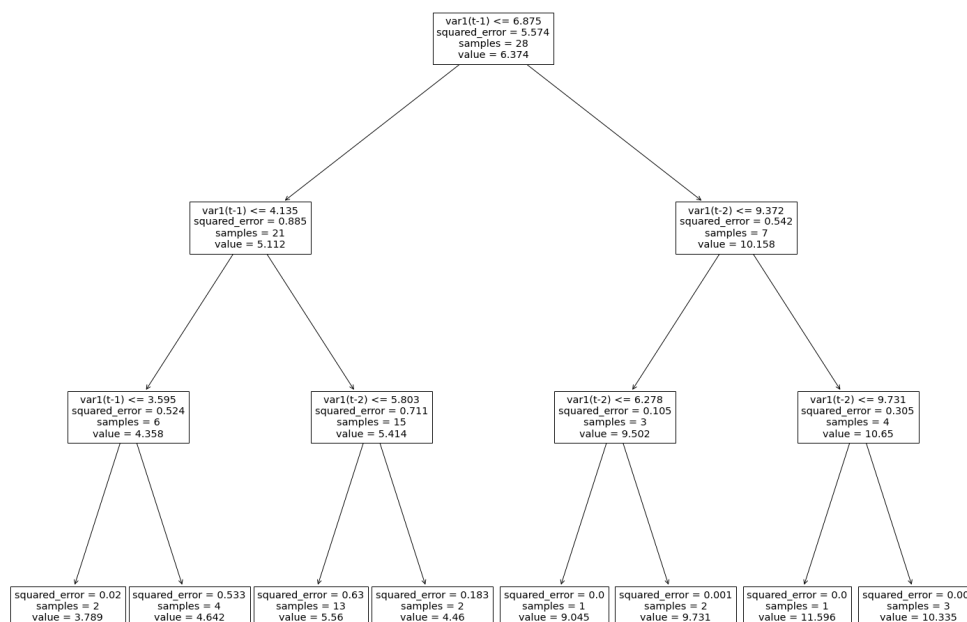


Figura 4 - Visualização de uma árvore de decisão³.

³ A primeira linha de cada caixa representa o critério de divisão para os próximos nós. A terceira linha, “samples”, informa o número de observações que se encaixam na categoria de cada um. A segunda, de “squared_error”, indica o erro quadrático médio quando se utiliza a última linha, “value”, como valor previsto em relação às demais observações do nó.

Em um problema de regressão, a variável resposta y_t explicada por X_1 e X_2 com oito retângulos (regiões) é descrita pela equação 14, onde c é uma constante e R uma região.

$$\hat{f}(X) = \sum_{m=1}^8 (c_m I\{(X_1, X_2) \in R_m\}) \quad (14)$$

Para construir uma árvore de decisão utilizando a soma do quadrado dos resíduos como critério de minimização e a média de y_i na região R_m como melhor \hat{c}_m , a melhor partição pode ser definida inicialmente pela definição do par de meios-planos na equação 15. A variável divisória j e o ponto de divisão s são obtidos pela resolução da equação 16. Depois de encontrar a melhor divisão, o processo é repetido para cada uma das sub-regiões até se encontrar o menor erro possível.

$$R_1(j, s) = \{X|X_j \leq s\} \text{ e } R_2(j, s) = \{X|X_j > s\} \quad (15)$$

$$\min_{j,s} \left[\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right] \quad (16)$$

No entanto, um modelo sobreajustado em seu treinamento, isto é, uma árvore complexa e com maior profundidade, pode resultar em maiores erros na projeção fora de amostra. Para contornar esse problema, considera-se o tamanho de árvore como um parâmetro de complexidade do modelo (HASTIE *et al*, 2009): a subdivisão em novos nós, por exemplo, só poderia ocorrer quando a queda na soma do quadrado dos resíduos em decorrente da divisão ultrapassasse algum limite preestabelecido. Também é possível impor um limite máximo de divisões à árvore para depois ajustá-la com um critério de complexidade de custo

$$C_\alpha(T) = \sum_{m=1}^{|T|} N_m Q_m(T) + \alpha |T|, \quad (17)$$

onde T é uma subárvore, $|T|$ é o número de nós terminais em T e m é um nó terminal da região R_m , critério por meio do qual se busca obter para cada α a árvore que minimiza $C_\alpha(T)$ unificando quaisquer números de nós internos não terminais.

3.4.1 Boosting

Árvores de decisão individuais podem ter as predições agregadas por meio de *bagging* ou *boosting*, gerando um modelo final pela técnica de *ensemble* que tende a menores erros

(WADE, 2020). Enquanto algoritmos de *bagging* utilizam o agregado de árvores de decisão, os de *boosting* se valem dos erros de árvores individuais, ou seja, se ajustam com base no erro de árvores anteriores e transformam vários *weak learners*, que têm uma taxa de erro levemente superior à escolha aleatória, em um *strong learner*. Inicialmente aplicado pelo algoritmo Adaptive Boosting (FREUND; SCHAPIRE, 1997), foi atualizado por Friedman (2001) no que ficou conhecido como Gradient Boosting Machine, cujo objetivo é minimizar a função custo pela adição de *weak learners* (nesse caso, árvores de decisão sem grande profundidade) por meio de procedimentos com gradiente descendente. Kuhn e Johnson (2016) expõem o funcionamento de uma versão simples desse algoritmo em sete passos:

1. Seleciona-se o tamanho D da árvore e o número K de iterações
2. Obtém-se a resposta média \bar{y} , utilizada como valor predito inicial de cada amostra
3. De k=1 a K
 - a. Calcula-se o resíduo de cada amostra
 - b. Estima-se uma árvore de decisão de tamanho D usando os resíduos como resposta
 - c. Prevê-se cada amostra pela estimação da árvore no passo anterior
 - d. Atualiza-se o valor previsto de cada amostra pela adição do valor previsto na iteração anterior ao valor ao valor previsto gerado no passo anterior

Os autores ressaltam que essa estratégia é suscetível a um sobreajustamento do modelo de forma local, já que otimiza os parâmetros apenas para os dados disponíveis, e pode obter maiores erros quando aplicada a predições fora de amostra. Um incremento àqueles passos para evitar isso é a inserção de um termo de regularização no último estágio, de forma que, em vez de atualizar o valor previsto de cada amostra pela adição do previsto na interação anterior, é adicionada apenas uma fração λ com valores entre 0 e 1.

3.4.2 XGBoost

O modelo Extreme Gradient Boosting (XGBoost) é uma versão avançada de *gradient boosting* desenvolvida por Chen e Guestrin (2016) e utilizado em problemas de classificação e de regressão. Com a mesma estrutura geral de transformação de *weak learners* em *strong learners*, apresentou melhores resultados que outros métodos de aprendizado de máquina para

séries temporais em estudos mais recentes (NIELSEN, 2019). O algoritmo funciona pela construção aditiva de árvores de decisão, cada uma buscando prever os resíduos da combinação das árvores anteriores. No entanto, ela é restrita pela minimização de uma função custo que contém um termo de regularização para a complexidade do modelo e que limita o número de árvores produzidas, consequentemente impedindo o sobreajustamento com infinitos modelos. A função objetivo a ser otimizada pode ser visualizada na equação 18, composta por uma função custo $l(y_i, \hat{y}_i^{(t)})$, que agrega as árvores anteriores e a atual, e pelo termo de regularização $\omega(f_t)$, definido na equação 19, na qual γ é uma penalidade pré-definida, T é o número de nós terminais (folhas), e w é o vetor de pontuações nas *folhas*.

$$obj = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \sum_{i=1}^t \omega(f_i) \quad (18)$$

$$\omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (19)$$

Para encontrar o resultado ótimo para uma folha, o algoritmo utiliza um polinômio de Taylor de segunda ordem, chegando à função objetivo na equação 20, onde g representa a derivada da função custo (gradiente) e h a sua segunda derivada (matriz Hessiana). Ao se unificar a função custo com o termo de regularização, obtém-se

$$obj^{(t)} \approx \sum_{i=1}^n [g_i w_{q(x_i)} + \frac{1}{2} h_i w_{q(x_i)}^2] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (20.a)$$

$$= \sum_{j=1}^T [(\sum_{i \in I_j} g_i) w_j + \frac{1}{2} (\sum_{i \in I_j} h_i + \lambda) w_j^2] + \gamma T \quad (20.b)$$

Com isso, é possível calcular o peso ótimo w_j^* do nó terminal j e medir a qualidade das estruturas das árvores de decisão, determinadas por um *greedy algorithm*⁴, com a função de minimização número 21,

⁴ Estratégia algorítmica que busca a melhor estratégia de cada etapa local com o objetivo de alcançar uma solução de ótimo global.

$$\min obj = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T, \quad (21)$$

onde $G_j = \sum_{i \in I_j} g_i$ e $H_j = \sum_{i \in I_j} h_i$.

3.4.3 LightGBM

O método *Light Gradient Boosting Machine* (KE et al., 2017) foi desenvolvido com o objetivo de otimizar a implementação do algoritmo de *gradient boosting* em árvores de decisão e introduziu as noções de *Gradient-based One-Side Sampling* (GOSS) e de *Exclusive Feature Bundling* (EFB). A primeira é uma técnica de amostragem e modificação do método de *gradient boosting*, a qual utiliza os exemplos de treinamento com maior gradiente resultante por se tratar de observações com maior significância no ganho informacional, consequentemente permitem uma otimização do tempo de estimação e reduzem a complexidade computacional do modelo. Para a construção de um modelo de *gradient boosting* usual, a árvore de decisão divide cada nó pela variável explicativa mais informativa, o que é medido pela variância após a divisão:

$$V_{j|o}(d) = \frac{1}{n_o} \left(\frac{\left(\sum_{\{x_i \in O: x_{ij} \leq d\}} g_i \right)^2}{n_{l|o}^j(d)} + \frac{\left(\sum_{\{x_i \in O: x_{ij} > d\}} g_i \right)^2}{n_{r|o}^j(d)} \right) \quad (22)$$

No método GOSS proposto, os exemplos de treinamento são primeiramente classificados de acordo com o valor absoluto dos seus gradientes em ordem decrescente. Depois disso, são mantidas as $\alpha \times 100\%$ observações com maiores gradientes, obtendo assim um subgrupo A de exemplos. Para o conjunto restante $(1-\alpha) \times 100\%$, de menores gradientes, é realizada uma amostragem aleatória, formando o subgrupo B. Por fim, as observações são divididas de acordo com o ganho de variância estimado $\tilde{V}_j(d)$ sobre o conjunto $A \cup B$, como mostra a equação 23. A estimação apenas sobre um subgrupo, e não sobre todas observações, auxilia na eficiência do modelo frente àqueles como o XGBoost e ainda, de acordo com os autores, não sofre com uma perda significativa de acurácia no treinamento, além de ter melhores resultados que uma amostragem aleatória para determinar os pontos de divisão da árvore.

$$\tilde{V}_j(d) = \frac{1}{n} \left(\frac{\left(\sum_{x_i \in A_l} g_i + \frac{1-a}{b} \sum_{x_i \in B_l} g_i \right)^2}{n_l^j(d)} + \frac{\left(\sum_{x_i \in A_r} g_i + \frac{1-a}{b} \sum_{x_i \in B_r} g_i \right)^2}{n_r^j(d)} \right) \quad (23)$$

A técnica EFB atua como uma seleção automática de variáveis relevantes para o modelo, empacotando aquelas mutuamente exclusivas em uma só para fins de eficiência. Isso ocorre pela construção de um grafo com pesos associados às arestas, os quais correspondem ao total de conflitos entre variáveis explicativas. Estas são então classificadas e atribuídas a um grupo de baixo conflito, ou então a um novo grupo.

3.5 TBATS

O modelo TBATS (*Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend and Seasonal Components*), desenvolvido por De Livera, Hyndman e Snyder (2011) utiliza uma combinação de séries de Fourier com alisamento exponencial e uma transformação Box-Cox para heterogeneidade da série de forma automatizada para capturar múltiplos padrões de sazonalidade. O modelo tem a vantagem de permitir que a sazonalidade mude ao longo do tempo, mas é custoso computacionalmente para a estimação (ATHANASOPOULOS; HYNDMAN, 2021). Pode ser descrito como

$$y_t^{(\omega)} = \ell_{t-1} + \phi b_{t-1} + \sum_{i=1}^T s_{t-m_i}^{(i)} + d_t, \quad (24.a)$$

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha d_t, \quad (24.b)$$

$$b_t = (1 - \phi)b + \phi b_{t-1} + \beta d_t, \quad (24.c)$$

$$s_t^{(i)} = s_{t-m_i}^{(i)} + \gamma_i d_t, \quad (24.d)$$

$$d_t = \sum_{i=1}^p \varphi_i d_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t, \quad (24.e)$$

$$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin \lambda_j^{(i)} + \gamma_1^{(i)} d_t, \quad (24.f)$$

$$s_{j,t}^{*(i)} = -s_{j,t-1} \sin \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} d_t, \quad (24.g)$$

onde $i = 1, \dots, T$; d_t é um processo ARMA (p, q) ; α , β e γ_i são parâmetros de alisamento; m_1, \dots, m_T são os períodos sazonais; ℓ_t é o nível local no período t ; b é a tendência de longo prazo; b_t é a tendência de curto prazo no período t ; $s_t^{(i)}$ é o i -ésimo componente sazonal em t ; ε é um processo ruído branco com média zero e variância constante; $\gamma_1^{(i)}$ e $\gamma_2^{(i)}$ são parâmetros de alisamento e $\lambda_j^{(i)} = 2\pi j/m_i$.

3.6 ENSEMBLE LEARNING

Modelos de *ensemble learning* não raramente são utilizados na literatura de previsão de demanda elétrica, como pode ser visto nos trabalhos de Fan *et al* (2014) e Guo *et al* (2021). Trata-se da combinação de resultados dos modelos $f_i(x)$ estimados individualmente para se obter uma melhor performance nas projeções, como descrito na equação

$$\hat{f}(x) = \sum_{i=1}^K \omega_i f_i(x), \quad (25)$$

onde $\sum_{i=1}^K \omega_i = 1$. A vantagem desse modelo é a sua capacidade de generalização, usualmente gerando erros e variância menores que os componentes isolados do *ensemble* (WICHARD; OGORZALEK, 2004). Algumas estratégias para a definição dos modelos e do peso ω_i são a de estimar os modelos com base em diferentes amostras dos dados e a de selecionar os modelos representativos com base no agrupamento de resultados. Neste trabalho, serão inicialmente computadas as projeções dos modelos descritos acima para, depois, gerar todas as combinações possíveis entre elas pela média dos valores. Por fim, será selecionada aquela que apresentar menor MAPE para o período de teste de quinze dias.

4 DADOS, ESTRATÉGIAS DE PREVISÃO E ACURÁCIA

4.1 DADOS

Os dados utilizados para o treinamento dos algoritmos e a projeção de demanda elétrica envolvem as séries diárias de carga no subsistema Sul do SIN – que abrange os estados do Paraná, Santa Catarina e Rio Grande do Sul – de 01/01/2000 a 31/05/2022 e os dados meteorológicos disponibilizados pelo Instituto Nacional de Meteorologia (INMET) de 01/01/2000 a 31/05/2022 para 95 estações no sul do Brasil, entre os quais estão a precipitação total, a pressão atmosférica, a temperatura de bulbo seco e do ponto de orvalho, a umidade relativa do ar e a velocidade média do vento. A carga de energia elétrica, medida em *megawatts* médios, é aquela injetada nos sistemas de transmissão e distribuição para atender ao consumo de energia elétrica dos consumidores, a perdas internas e a possíveis diferenças na rede. Um *megawatt médio* é a energia média no intervalo de tempo considerado; no caso de estudo, a base é diária ($1 \text{ MW}_{\text{med-dia}} = 24 \text{ MWh/dia}$) e foi utilizada integralmente para a estimação dos modelos baseados em processos univariados, como SARIMA e Prophet. Para os multivariados, como XGBoost e LightGBM, que utilizam também as informações meteorológicas, os dados empregados iniciam em 01/01/2008. Embora os dados do INMET para antes de 2008 sejam disponibilizados, uma inconsistência verificada na base para esse período sugere que os algoritmos possam buscar informações mais assertivas a partir daquele ano. O uso de 15 defasagens de cada variável climática e da série de demanda se mostrou o mais eficiente em termos de MAPE e custo computacional, resultando em um total de 285 variáveis explicativas para as estratégias de *boosting*. Na figura 5, as séries podem ser visualizadas integralmente, com as meteorológicas representadas pela média diária nos três estados. Na estimação do modelo, foi considerada a média das variáveis mensuradas nas estações para cada estado, totalizando 18 variáveis climáticas.

Ao se observarem as figuras 10 e 11, no apêndice do trabalho, é possível verificar dois padrões de sazonalidade para a série de demanda elétrica. Um deles tem maior consumo nos meses do início do ano, quando há uma elevação da temperatura e, com isso, do uso de aparelhos movidos a energia utilizados em domicílios e ambientes de trabalho. O outro padrão é o de sazonalidade semanal, com uma queda no consumo aos finais de semana. Este segundo

é explicado em maior parte pela diferença no perfil de consumo de energia elétrica, já que não se desagrega a demanda por faixas de tensão, ou seja, tanto o consumo industrial quanto doméstico são contabilizados. Logo, com uma redução da atividade econômica aos sábados e domingos, também sofre uma queda a demanda por energia elétrica.

Ainda no apêndice do trabalho, a figura 12 indica uma amplificação do efeito sazonal na série com o passar dos anos, o que condiz com uma natureza multiplicativa da sazonalidade. Isso também é indicado para a sazonalidade semanal pelo desvio padrão com janela móvel de sete dias, que aumenta com o tempo.

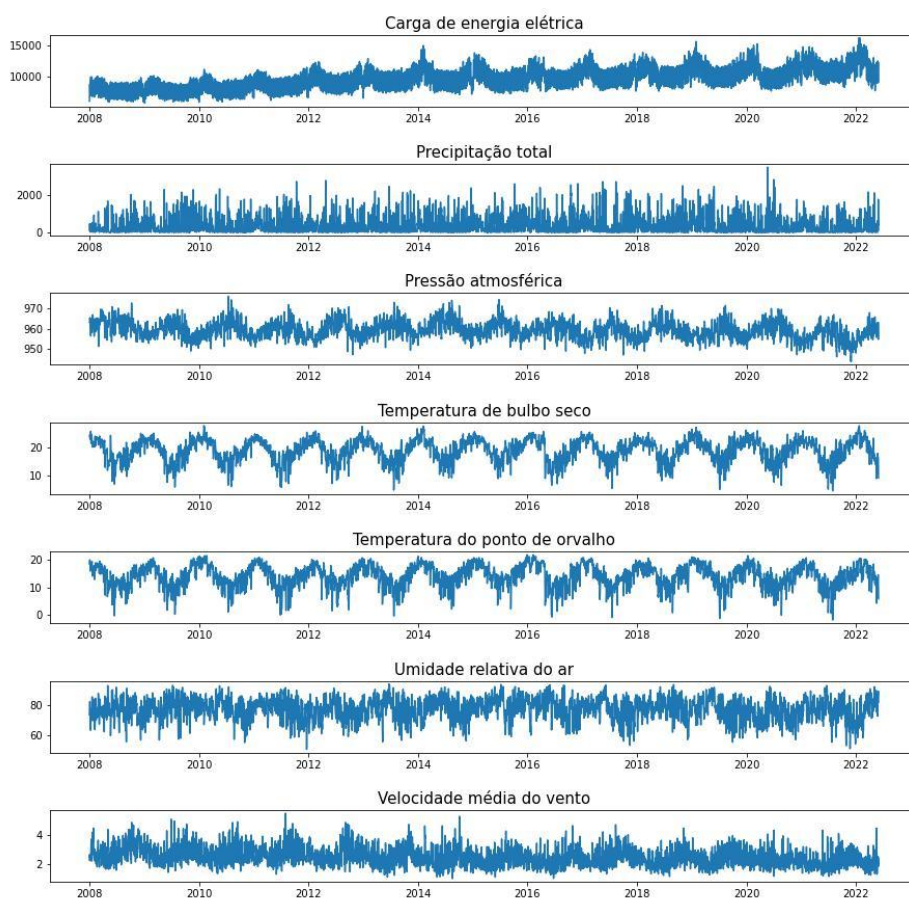


Figura 5 - Variáveis analisadas.

Observando o gráfico da função de autocorrelação na série de carga elétrica (figura 6), há indícios da presença de raiz unitária. Para verificar a condição de estacionariedade necessária a alguns dos modelos antes da estimação dos parâmetros, foram realizados os testes ADF e KPSS para a série em nível.

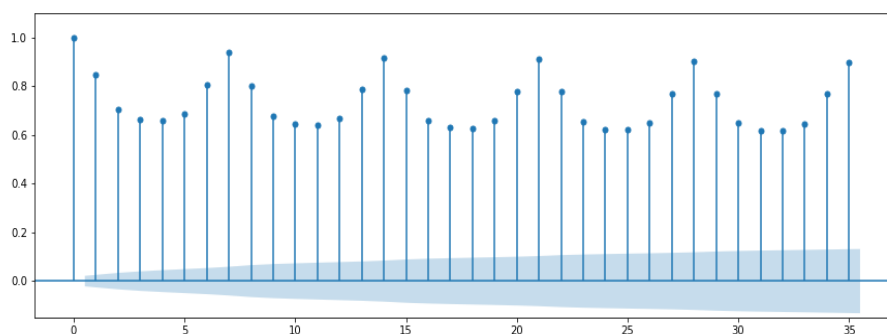


Figura 6 - FAC da série de carga elétrica.

O p-valor do teste KPSS, cuja hipótese nula é a de um processo estacionário, foi significativo (inferior a 0,01); o do ADF, com hipótese nula de presença de raiz unitária, não (0,1239). Ambos os testes sugerem estacionariedade da série quando esta passa por uma diferenciação de grau 1.

Uma análise mais extensiva dos dados climáticos e de demanda elétrica está disponível no repositório citado no capítulo 1.

4.2 ESTRATÉGIAS DE PREVISÃO

Existem dois cenários para projeção de uma série temporal. Um deles é aquele onde se faz necessária a projeção para apenas um passo à frente ($t+1$). O outro é aquele em que se busca projetar um horizonte com h períodos à frente ($t+h$) de forma sequencial. A diferenciação teórica e prática entre as duas foi fundamental, uma vez que foram estimados modelos univariados e multivariados até o período T . Como o intuito do trabalho foi garantir um horizonte mínimo de projeções confiáveis no curto prazo, a ênfase foi em projeções *multi-step*.

Bontempi *et al* (2013) expõem, dentro da classe de previsões sequenciais, dois métodos. O primeiro deles, utilizado neste trabalho para os modelos univariados, é a estratégia recursiva, na qual se estima um único modelo para o valor em $t+1$ e, a partir deste, é realizada uma projeção com múltiplos passos recursivamente. Um ponto negativo desse método é que, por realizar uma projeção a partir de projeções anteriores, pode carregar os erros iniciais progressivamente. Outro, que, para realizar uma projeção dos modelos multivariados no contexto do presente estudo, seriam necessárias projeções confiáveis das variáveis climáticas regressoras para toda a região sul até $t + h$ sendo geradas simultaneamente. O segundo

método, utilizado tanto para os univariados quanto para os multivariados, é a estratégia direta, que estima um modelo para cada período à frente da projeção, ou seja, para um horizonte de 7 dias, seriam treinados 7 modelos independentes, cujos resultados são por fim agrupados em sequência. Esse método não apresenta a mesma problemática que a estratégia recursiva, pois não acumula o erro de projeções anteriores, mas, por estimar modelos independentes, não considera dependências entre as previsões. Outro ponto negativo dessa abordagem é o custo computacional, porque ao contrário da recursiva, que treina apenas um modelo, esta estima tantos modelos quanto maior for o horizonte de projeção.

Os algoritmos de árvores de decisão, XGBoost e LightGBM, geralmente utilizados em problemas de classificação e regressão, foram aplicados à série de demanda elétrica transformando as defasagens em variáveis explicativas, como ilustrado tabela 1, retirando as linhas que continham valores nulos. Como explicado acima, a abordagem univariada permite que os três modelos sejam estimados de forma recursiva e direta.

Data	Yt	Yt-1	Yt-2
01/01/2000	4.800,65		
02/01/2000	4.899,80	4.800,65	
03/01/2000	6.261,55	4.899,80	4.800,65
04/01/2000	6.733,74	6.261,55	4.899,80
05/01/2000	6.961,17	6.733,74	6.261,55
06/01/2000	7.110,36	6.961,17	6.733,74
07/01/2000	7.105,35	7.110,36	6.961,17
08/01/2000	6.307,49	7.105,35	7.110,36
09/01/2000	5.523,62	6.307,49	7.105,35
10/01/2000	7.111,32	5.523,62	6.307,49
11/01/2000	7.435,06	7.111,32	5.523,62
12/01/2000	7.425,49	7.435,06	7.111,32

Tabela 1: Transformação da série em um problema de regressão.

Para a abordagem multivariada, por outro lado, foi utilizada apenas a forma direta. O tratamento de dados segue a mesma lógica que a anterior, com a diferença de que se substitui Y_t por $Y_{t+1}, Y_{t+2}, \dots, Y_{t+h-1}$ para cada um dos modelos destinados a prever $t+h$.

4.3 ACURÁCIA

Para verificar a performance dos modelos, a série foi inicialmente dividida em uma seção de treino, de 01/01/2000 a 30/04/2022 para os modelos univariados e de 01/04/2006 a 30/04/2022 para os multivariados, utilizada para treinar os parâmetros; e outra de teste, que

abrange o período de 01/05/2022 a 31/05/2022. Com esta, é possível realizar uma avaliação das projeções fora de amostra entre a observação Y_i e o seu valor projetado \hat{Y}_i no tempo t . Algumas das medidas de acurácia mais frequentemente utilizadas são o Erro Médio Absoluto (do inglês *Mean Absolute Error*, MAE), Erro Quadrático Médio (*Mean Squared Error*, MSE), Raiz do Erro Quadrático Médio (*Root Mean Squared Error*, RMSE), Erro Percentual Absoluto Médio (*Mean Absolute Percentage Error*, MAPE) e Erro Percentual Absoluto Médio Simétrico (*Symmetric Mean Absolute Percentage Error*, sMAPE), cujas equações são descritas na tabela 2.

<p>Erro Médio Absoluto (MAE)</p> $\left(\frac{1}{n}\right) \sum_{t=1}^n Y_t - F_t $
<p>Erro Quadrático Médio (MSE)</p> $\left(\frac{1}{n}\right) \sum_{t=1}^n (Y_t - F_t)^2$
<p>Root Mean Squared Error (RMSE)</p> $\sqrt{\left(\frac{1}{n}\right) \sum_{t=1}^n (Y_t - F_t)^2}$
<p>Erro Percentual Absoluto Médio (MAPE)</p> $\left(\frac{1}{n}\right) \sum_{t=1}^n \frac{ Y_t - F_t }{Y_t}$
<p>Erro Percentual Absoluto Médio Simétrico (sMAPE)</p> $\left(\frac{200}{n}\right) \sum_{t=1}^n \frac{ Y_t - F_t }{ Y_t + F_t }$

Tabela 2: Medidas de acurácia.

As medidas MAE, MSE e RMSE dependem da escala da série analisada, não sendo possível compará-las em processos medidos em diferentes unidades. Escolher o modelo com menor MAE levará a projeções da mediana; com menor RMSE, da média (ATHANASOPOULOS; HYNDMAN, 2021), também atribuindo, junto do MSE, maiores pesos à presença de *outliers*. O MAPE tem como ponto positivo a independência da unidade de medida da série e é frequentemente utilizado na comparação entre conjuntos de dados

diferentes, além de apresentar maior interpretabilidade em relação ao restante dos métodos. No entanto, entre alguns problemas do MAPE estão o fato de distorcer as comparações no caso de *outliers* e gerar maiores erros percentuais absolutos para desvios negativos e baixos valores observados. Estas distorções instigaram o uso do sMAPE, que, diferentemente de sua primeira versão, estabelece um limite inferior e um limite superior de 0% e 200%, respectivamente. A escolha de analisar múltiplas medidas de acurácia advém do fato de que, do ponto de vista teórico, não existe o “melhor” método, mas aquele relacionado ao contexto do processo (MAKRIDAKIS, 1993), de modo que os resultados agregados permitem uma melhor interpretação das projeções.

5 RESULTADOS

Nesta seção são avaliadas as projeções realizadas para o período de quinze dias entre 17/05/2022 e 31/05/2022. Foram comparados os modelos de Holt-Winters, SARIMA, TBATS, Prophet, árvores de decisão, LightGBM e XGBoost univariados com as estratégias de projeção recursiva e direta, LightGBM e XGBoost multivariados com estratégia direta e a combinação linear das projeções com menor MAPE (ensemble), descritos no capítulo 3. Na tabela 3 constam as métricas apresentadas na seção 4.1 para cada um destes modelos em relação aos valores observados nos quinze dias, comparação que pode ser verificada na figura 13 do Apêndice A. Na figura 7 são apresentados gráficos de *boxplot* com os erros absolutos percentuais (APE) em cada projeção do horizonte. Por fim, na tabela 4 é exposto o MAPE dos modelos para cada período entre $t+1$ e $t+15$.

Modelo	MAE	MSE	RMSE	MAPE	sMAPE
Árvores de decisão	706,88	1.022.618,46	1.011,25	6,20%	6,57%
Ensemble	193,22	73.986,02	272,00	1,60%	1,63%
Holt Winters	405,41	257.589,52	507,53	3,50%	3,60%
LightGBM multivariado	355,93	189.093,18	434,85	3,07%	3,12%
LightGBM univ. (direto)	595,94	461.835,64	679,58	5,20%	5,37%
LightGBM univ. (recursivo)	522,42	316.555,55	562,63	4,63%	4,76%
Prophet	265,05	112.093,07	334,80	2,50%	2,46%
SARIMA	601,10	453.907,98	673,73	5,19%	5,36%
TBATS	798,20	915.860,54	957,01	6,77%	7,07%
XGBoost multivariado	517,14	385.797,12	621,13	4,43%	4,55%
XGBoost univ. (direto)	682,64	653.348,19	808,30	5,90%	6,15%
XGBoost univ. (recursivo)	474,41	278.864,66	528,08	4,16%	4,27%

Tabela 3: Acurácia dos principais modelos⁵.

O modelo SARIMA estimado com menor AIC foi o de ordem (1, 1, 2)(1,0,1)₇ com tendência para a série logaritimizada, pois esta transformação apresentou menor MAPE quando comparada à série original. Sua projeção ficou em oitavo lugar de acordo com todas as métricas utilizadas, à exceção do erro absoluto médio, tendo desempenho ainda inferior ao

⁵ A tabela está classificada com um gradiente de cor para cada coluna, onde a escala de verde equivale a menores erros dos modelos; a de vermelho, a maiores.

do modelo LightGBM univariado com estratégia de projeção recursiva. O modelo de Holt-Winters estimado com sazonalidade multiplicativa foi um dos melhores colocados, com acurácia superior às do XGBoost univariado recursivo, XGBoost multivariado e LightGBM univariado recursivo. Com quinze defasagens das variáveis climáticas e da série de demanda elétrica, o LightGBM multivariado ficou em terceiro lugar, com 3,07% de MAPE, evidenciando um ganho informacional com a inclusão das variáveis explicativas. Em segundo lugar ficou o algoritmo Prophet, com 2,50% de MAPE, mantendo a posição também pelas outras métricas. O modelo com menor MAPE foi o de *ensemble*, gerado a partir da combinação linear entre as projeções do Prophet e Holt-Winters, com 1,60% de erro percentual absoluto médio. Os modelos de *boosting* univariados com estratégia de projeção recursiva apresentaram um desempenho superior àqueles de projeção direta.

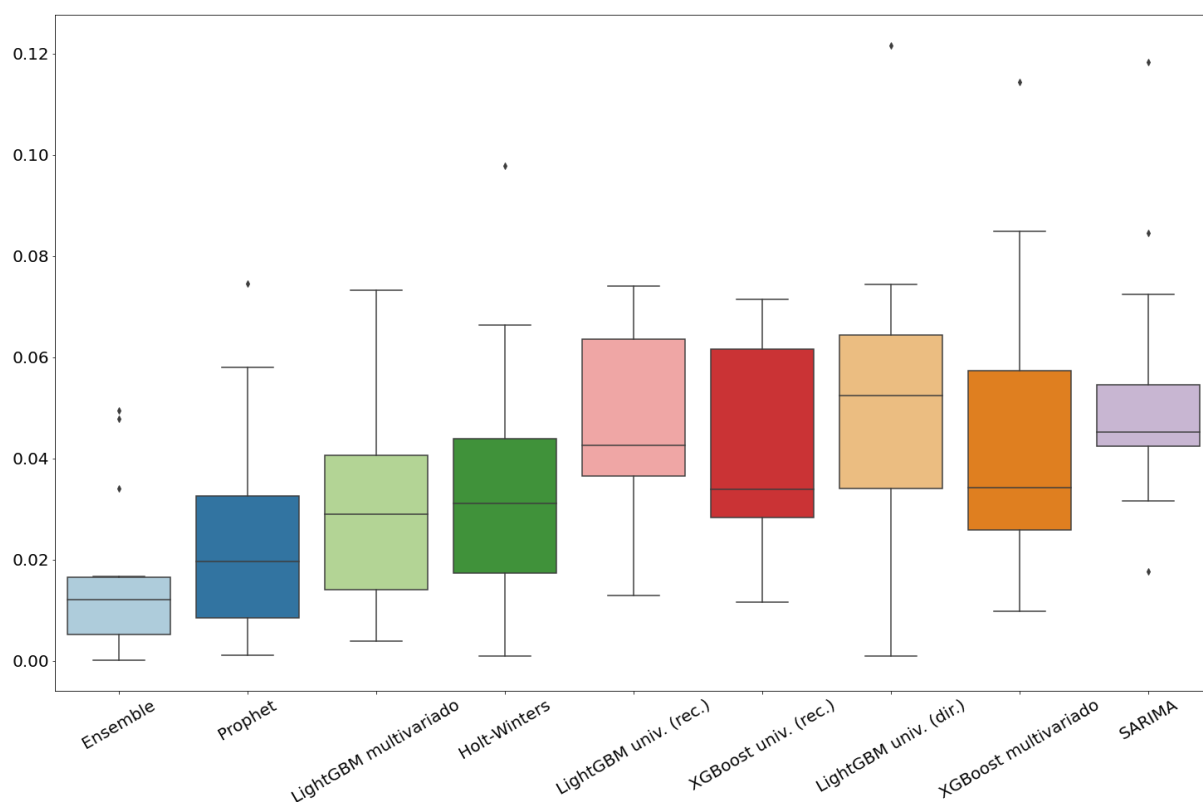


Figura 7: APE dos melhores modelos.

As distribuições dos erros de cada modelo foram dispostas na figura 7 e ordenadas no eixo x pelo percentil 90 dos erros percentuais absolutos para cada t . O modelo SARIMA, embora apresente três erros percentuais absolutos (APE) como *outliers*, é um daqueles com menor variância dos erros, o que indica uma estabilidade do modelo no horizonte de projeção. A abordagem de Holt-Winters, que havia sido uma das melhores colocadas pelo MAPE,

gerou erros mais dispersos. O *ensemble* se mostra novamente como o mais acurado, com os menores números de APE e menor variância dos erros, apesar dos três valores anormais.

Modelo	t+1	t+2	t+3	t+4	t+5	t+6	t+7	t+8	t+9	t+10	t+11	t+12	t+13	t+14	t+15
SARIMA	1,77%	3,04%	3,42%	4,01%	4,10%	3,95%	4,01%	3,94%	4,00%	4,45%	5,12%	5,30%	5,24%	5,19%	5,19%
Árvores de decisão	2,31%	1,85%	5,86%	5,25%	5,98%	6,08%	5,37%	4,86%	4,74%	4,38%	4,54%	4,76%	4,78%	4,88%	6,20%
Ensemble	0,79%	1,08%	0,78%	0,72%	0,69%	0,82%	0,79%	0,76%	0,82%	1,18%	1,68%	1,75%	1,67%	1,58%	1,60%
Holt Winters	0,38%	1,51%	1,66%	2,14%	2,64%	2,95%	2,83%	2,71%	2,71%	3,07%	3,68%	4,04%	4,26%	4,13%	4,08%
LightGBM multivariado	1,86%	3,04%	2,99%	2,34%	2,20%	2,72%	2,42%	2,26%	2,72%	2,84%	3,25%	3,14%	2,97%	3,03%	3,07%
XGBoost multivariado	0,97%	3,47%	3,41%	3,37%	3,06%	3,29%	3,18%	3,84%	4,02%	4,77%	4,93%	4,71%	4,64%	4,55%	4,43%
Prophet	1,95%	1,03%	1,22%	1,53%	2,39%	3,23%	2,90%	2,63%	2,36%	2,37%	2,49%	2,52%	2,76%	2,66%	2,50%
TBATS	9,69%	7,84%	5,69%	5,44%	5,84%	5,31%	6,78%	6,83%	6,68%	6,87%	7,04%	6,53%	6,05%	6,69%	6,77%
LightGBM univ. - direto	0,09%	2,52%	2,71%	2,72%	3,05%	3,14%	3,74%	3,95%	4,10%	4,90%	5,02%	5,22%	5,33%	5,34%	5,20%
LightGBM univ. - recursivo	1,29%	4,00%	3,99%	4,21%	4,62%	4,93%	4,83%	4,73%	4,57%	4,86%	4,95%	4,88%	5,02%	4,76%	4,63%
XGBoost univ. - direto	0,12%	3,68%	3,81%	3,27%	3,01%	3,92%	4,51%	5,03%	5,29%	5,36%	6,24%	6,12%	5,99%	5,95%	5,90%
XGBoost univ. - recursivo	1,40%	2,88%	2,94%	3,06%	3,00%	3,07%	3,03%	2,79%	2,80%	3,24%	3,58%	3,86%	4,05%	4,02%	4,16%

Tabela 4: MAPE em $t+h^6$.

Pela análise cumulativa de MAPE das projeções no horizonte de quinze dias, a técnica de *ensemble* foi consistentemente a mais acurada entre $t+3$ e $t+15$, mantendo a métrica inferior a 1,08% até $t+9$ e atingindo um mínimo de 0,69% de erro percentual absoluto médio para uma projeção até $t+5$. Para a projeção um dia à frente, o modelo LightGBM univariado por estratégia direta foi o de menor erro, com 0,09%, seguido pelo XGBoost univariado por estratégia direta, com 0,12%. As técnicas de *boosting* com estratégia direta apresentaram menor erro nas projeções de menor prazo, de um a cinco dias, mas suas performances deterioraram no restante do período enquanto as de estratégia recursiva melhoraram sua acurácia. Um dos possíveis motivos para tanto é a perda informacional das predições subsequentes que ocorre nos modelos projetados de forma direta, conforme apontado na seção 4.2. O TBATS foi o pior colocado em quase todas as projeções e gera um dos custos computacionais mais elevados entre as abordagens desenvolvidas neste trabalho, o que coloca em dúvida sua aplicabilidade no contexto de um mercado de curto prazo. A projeção com alisamento exponencial de Holt-Winters manteve-se com um MAPE relativamente baixo em todo o período e ganha destaque por ser um modelo tradicional relativamente simples quando comparado aos demais métodos.

⁶ De forma semelhante à da tabela 3, os gradientes de cores estão relacionados ao MAPE de cada coluna.

5.1 BACKTESTING

O *cross-validation* é um método de amostragem utilizado para verificar a capacidade de generalização de modelos preditivos, para prevenir sobreajustamento e para comparar algoritmos de aprendizagem de máquina (LIU; OZSU, 2009). Em um *k-fold cross-validation*, os dados são repartidos em k segmentos aleatórios de mesmo tamanho para depois serem realizadas k iterações subsequentes de estimação e validação, de modo que a cada iteração é deixada uma partição de teste para as $k-1$ partições de treino restantes. No entanto, quando se consideram análises de séries temporais, o procedimento não pode ser realizado da mesma forma, uma vez que elas apresentam autocorrelação serial e obedecem a uma ordem cronológica. É então que se utiliza o *backtesting*, método pelo qual se deslocam progressivamente as partições de treino e teste dos modelos até as observações mais recentes, em vez de fazê-lo aleatoriamente. Este processo está ilustrado na figura 8:

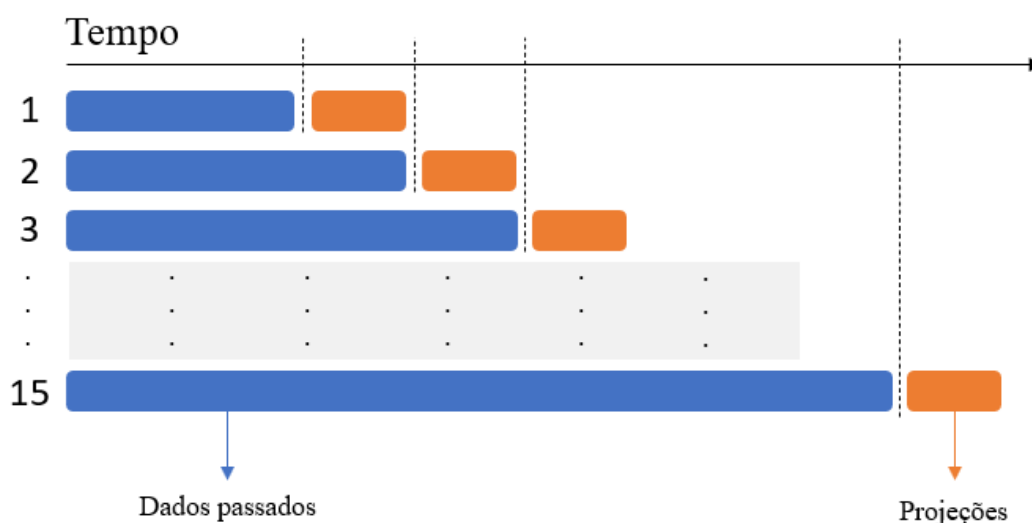


Figura 8: *Backtesting*.

Para verificar a representatividade das medidas de acurácia e da classificação dos modelos obtidas na seção de resultados, os quatro métodos com menor MAPE foram treinados em quinze partições sequenciais com um horizonte de projeção de quinze dias em cada, contemplando uma janela crescente de dados de entrada que começa em 04/10/2021 e termina em 16/05/2022 – ou seja, a última partição de teste é a mesma utilizada como referência para os resultados mais acima. Após computar as projeções, a distribuição dos erros

percentuais absolutos de cada modelo foi disposta na figura 9, onde os modelos estão classificados em ordem crescente de MAPE no eixo x .

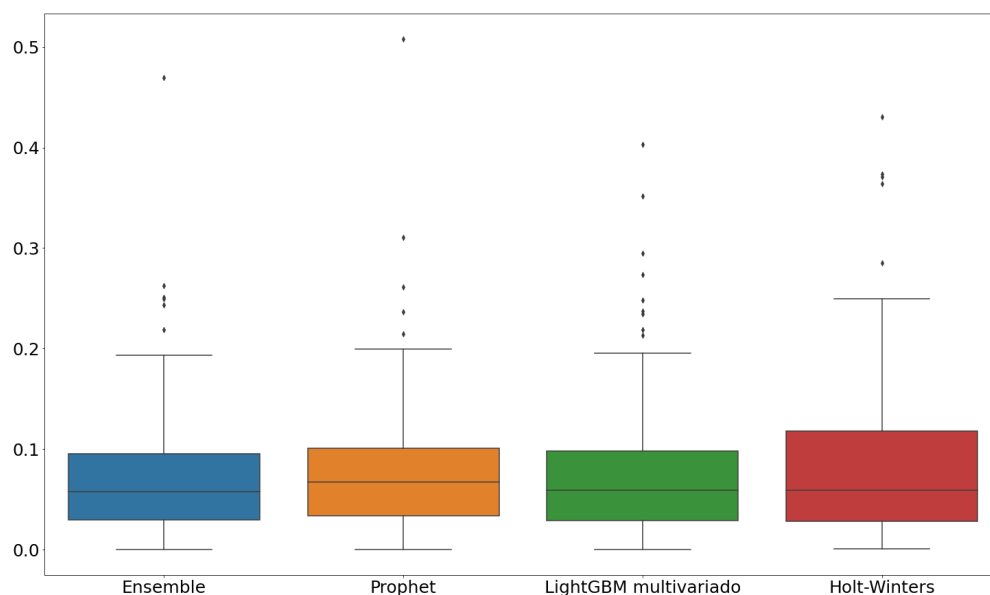


Figura 9: Distribuição dos erros em *backtesting*.

Embora seja possível verificar um maior desvio das projeções em relação aos valores reais, o processo de validação indicou consistência do que foi observado na subseção anterior: o *ensemble* foi o modelo de menor erro, com 6,83% de MAPE; o Prophet, o segundo, com 7,39%; o LightGBM multivariado, o terceiro, com 7,48%; e o Holt-Winters, o quarto, com 8,13%. Os três primeiros tiveram pelo menos 75% de seus erros percentuais absolutos inferiores a 10%, enquanto os do último apresentaram uma maior dispersão e valor médio.

6 CONSIDERAÇÕES FINAIS

Considerando modelos-base e estratégias de projeção, foram comparados ao todo 12 modelos em um horizonte de quinze dias com as medidas de acurácia *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE); *Root Mean Squared Error* (RMSE), *Mean Absolute Percentage Error* (MAPE) – a principal analisada – e *Symmetric Mean Absolute Percentage Error* (sMAPE). O modelo SARIMA com menor AIC foi o de ordem $(1, 1, 2)(1,0,1)_7$ para a série logaritmicada e obteve um MAPE de 5,19% no período. A projeção com Holt-Winters foi melhor, com 3,50%. Como variáveis explicativas para os modelos multivariados, foram selecionadas quinze defasagens das médias de precipitação total, pressão atmosférica, temperatura de bulbo seco e do ponto de orvalho, umidade relativa do ar e velocidade média do vento por dia para cada um dos estados de Santa Catarina, Paraná e Rio Grande do Sul. Essa estruturação dos dados gerou um MAPE de 3,07% com o modelo LightGBM. O algoritmo Prophet, univariado, teve o melhor desempenho individual, com 2,49% de erro nos quinze dias. O modelo de *ensemble*, obtido pela média das projeções daquele com as de Holt-Winters, foi o que apresentou melhores resultados em todas as métricas utilizadas com 1,60% de MAPE.

Foi possível verificar uma superioridade dos modelos de aprendizado de máquina em relação às metodologias tradicionais de Holt-Winter e SARIMA. Essa comparação também foi reforçada pelo processo de *backtesting* para os quatro modelos com menor MAPE. Não obstante, o modelo de Holt-Winters se mostra como uma alternativa viável à previsão da série quando se leva em conta sua fácil interpretação para um método com 3,50% de erro percentual absoluto médio no período avaliado. O resultado geral corrobora, em certa medida, o que foi obtido por Massaoudi *et al* (2021) na utilização do algoritmo LightGBM para projeção de demanda elétrica, assim como o uso recorrente de técnicas de *ensemble* apresentado no capítulo referente à revisão de literatura. No contexto de demanda, o trabalho carece de dados referentes ao preço da energia elétrica, fator ressaltado por Andrade (2010) como um dos mais relevantes na projeção de curto prazo, e de maior granularidade dos dados de carga, como nível de tensão e unidades de mensuração mais segregadas, bastante presente na literatura revisada (ALTRAN, 2010; ANDRADE, 2010; UNSIHUAY-VILA *et al*, 2010; FAN *et al*, 2014; VELASCO *et al*, 2018; ZHENG *et al*, 2019; MAEDA, 2021). Não obstante, as projeções foram consideradas satisfatórias quando analisadas por critérios como o MAPE, pois os três principais modelos apresentaram erros inferiores a 5%.

REFERÊNCIAS

ANDRADE, L. C. M. **Abordagem Neurofuzzy para Previsão de Demanda de Energia Elétrica no Curtíssimo Prazo**. Dissertação (Mestrado) – Escola de Engenharia de São Carlos – Universidade de São Paulo, Programa de Engenharia Elétrica, São Carlos, 2010.

BASHIR, Z. A., EL-HAWARY, M. E. Applying Wavelets to Short-Term Load Forecasting Using PSO-Based Neural Networks. **IEEE Transactions on Power Systems**, v. 24, p. 20-27, 2009.

BRASIL, EMPRESA DE PESQUISA ELÉTRICA. **Anuário Estatístico de Energia Elétrica**. Rio de Janeiro, 2021.

BRASIL, EMPRESA DE PESQUISA ELÉTRICA. Metodologia: Modelo de Projeção da Demanda de Eletricidade. **Nota técnica DEA 003/19**. Rio de Janeiro, p. 1 – 33, 2019.

BRASIL, EMPRESA DE PESQUISA ELÉTRICA. Projeção de demanda de energia elétrica para os próximos 10 anos. **Nota técnica DEA 001/17**. Rio de Janeiro, p. 1 – 95, 2017.

BRASIL. Lei nº 10.848, de 15 de março de 2004. **Dispõe sobre a comercialização de energia elétrica, altera as Leis nºs 5.655, de 20 de maio de 1971, 8.631, de 4 de março de 1993, 9.074, de 7 de julho de 1995, 9.427, de 26 de dezembro de 1996, 9.478, de 6 de agosto de 1997, 9.648, de 27 de maio de 1998, 9.991, de 24 de julho de 2000, 10.438, de 26 de abril de 2002, e dá outras providências**. Diário Oficial da União, 2004. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/lei/110.848.htm. Acesso em: 17 set. 2022.

BENTÉJAC, Candice *et al.* **A comparative Analysis of XGBoost**. arXiv:1911.01914, 2019.

BONTEMPI, Gianluca *et al.* Machine Learning Strategies for Time Series Forecasting. **Lecture Notes In Business Information Processing**, v. 138, p. 62-77, 2013.

BORDIGNON, Sérgio. **Metodologia para previsão de carga de curtíssimo prazo considerando variáveis climáticas e auxiliando na programação de despacho de pequenas centrais hidrelétricas**. Dissertação (Mestrado) – Universidade Federal de Pampa, Programa de Pós-Graduação em Engenharia Elétrica, 2012.

BOX, George, JENKINS, Gwilym. **Time Series Analysis: Forecasting and control**. 1st ed. San Francisco: Holden-Day, 1970.

BOX, G. E., PIERCE, D. A. Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. **Journal of the American Statistical Association**, v. 65, n. 332, p. 1509–1526, 1970.

BRAGA, Lucas de Almeida. **Um estudo sobre o mercado de energia elétrica no Brasil**. Dissertação – Universidade Federal de Ouro Preto, Colegiado do Curso de Engenharia de Controle e Automação, Ouro Preto, 2018.

BUENO, Rodrigo. **Econometria de Séries Temporais**. 2ª edição. São Paulo: Cengage Learning, 2018.

CAMPOS, Roger Júnio. **Previsão de séries temporais com aplicações a série de consumo de energia elétrica**. Dissertação (Mestrado) – Universidade Federal de Minas Gerais, Colegiado do Programa de Pós-Graduação em Energia Elétrica, Belo Horizonte, 2008.

CERQUEIRA, Vitor *et al.* Evaluating time series forecasting models: An empirical study on performance estimation methods. **Machine Learning**, v. 109, n.11, p. 1–32, 2020.

CHEN, Tianqi; GUESTRIN, Carlos. XGBoost: A scalable tree boosting system. **International Conference on Knowledge, Discovery and Data Mining**, p. 785–794, New York, 2016.

CHEN, K. *et al.* Short-term Load Forecasting with Deep Residual Networks. **IEEE Transactions on Smart Grid**, v. 10, n. 4, p. 3943–3952, 2019.

DE LIVERA, A. M. *et al.* A State Space Framework for Automatic Forecasting Using Exponential Smoothing Methods. **Journal of the American Statistical Association**, v. 106, p. 1513–1527, 2011.

DIAS, Eduardo. **Previsão de Médio Prazo do Consumo de Energia Elétrica no Brasil: Estimção via Metodologia Box & Jenkins e Regressão Dinâmica**. Dissertação (Mestrado em Economia Aplicada) – Universidade Federal de Juiz de Fora, Programa de Pós-graduação em Economia Aplicada, Juiz de Fora, 2008.

DICKEY, D. A., FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. **Journal of the American Statistical Association**, v. 74, n. 366, p. 427–431, 1979.

DICKEY, D. A., FULLER, W. A. Likelihood ratio statistics for autoregressive time series with a unit root. **Econometrica**, v. 49, n. 4, p. 1057–1073, 1981

ENDERS, Walter. **Applied econometric time series**. 4th ed. New York: John Wiley & Sons, 2014.

FAWZY, Haitham *et al.* Time Series Forecasting Using Tree Based Methods. **Journal of Statistics Applications & Probability**, v. 10, n. 1, p. 229-244, 2021.

FRIEDMAN, Jerome. Greedy function approximation: A gradient boosting machine. **The Annals of Statistics**, v. 29, n. 5, p. 1189–1232, 2001.

GÉRON, Aurélien. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow**. California: O'Reilly Media, Inc. 2019.

GOODFELLOW, Ian *et al.* **Deep Learning**. Cambridge: MIT Press, 2017.

GRIGSBY, Jake *et al.* **Long-Range Transformers for Dynamic Spatiotemporal Forecasting**. arXiv:2109.12218, 2021.

GUJARATI, Damodar *et al.* **Econometria Básica**. 5ª edição. Porto Alegre: AMGH, 2011.

GUO, Weilin *et al.* Machine-Learning based methods in short-term load forecasting. **The Electricity Journal**, v. 34, n. 1, 2021.

HASTIE, Trevor *et al.* **An Introduction to Statistical Learning**. New York: Springer, 2021.

HASTIE, Trevor *et al.* **The Elements of Statistical Learning: Data Mining, Inference and Prediction**. 2nd ed. New York: Springer, 2009.

HIBON, Michele *et al.* Generalising about univariate forecasting methods: further empirical evidence. **International Journal of Forecasting**, v. 14, n. 3, p. 339 – 358, 1998.

HYNDMAN, Rob; ATHANASOPOULOS, George. **Forecasting: Principles and Practice**. 3rd ed. Melbourne: OTexts, 2021.

KAZEMI, Seyed *et al.* **Time2Vec: Learning a Vector Representation of Time**. arXiv: 1907.05321, 2019.

KE, Guolin *et al.* Lightgbm: a highly efficient gradient boosting decision tree. In: **Advances in neural information processing systems**, p. 3146–3154, 2017.

KHOTANZAD, Alireza *et al.* ANNSTLF -Artificial Neural Network Short-Term Load Forecaster - generation three. In: **IEEE Transactions on Power Systems**, v. 13, n. 4, p. 1413–1422, 1998.

KUHN, Max, JOHNSON, Kjell. **Applied Predictive Modeling**. New York: Springer, 2016.

KWIATKOWSKI, Denis *et al.* Testing the null hypothesis of stationarity against the alternative of a unit root. **Journal of Econometrics**, v. 54, p. 159–178, 1992.

LIM, Bryan; ZOHREN, Stefan. Time Series Forecasting With Deep Learning: A Survey. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 379, n. 2194, 2021.

LIM, Bryan *et al.* **Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting**. arXiv: 1912.09363, 2020.

LIU, Ling, ÖZSU, Tamer. **Encyclopedia of Database Systems**. New York: Springer, 2009.

LJUNG, G. M., BOX, G. E. "On a Measure of a Lack of Fit in Time Series Models". **Biometrika**, v. 65, n. 2, p. 297–303, 1978.

MAKRIDAKIS, Spyros. Accuracy measures: theoretical and practical concerns. **International Journal of Forecasting**, v. 9, n. 4, p. 527–529, 1993.

MAKRIDAKIS, Spyros *et al.* Statistical and Machine Learning forecasting methods: Concerns and ways forward. **PLOS ONE**, v. 13, n. 3, p. 1 – 26, 2018.

MAKRIDAKIS, Spyros *et al.* The M3-Competition: results, conclusions and implications. **International Journal of Forecasting**, v. 16, n. 4, p. 451 – 476, 2000.

MAKRIDAKIS, Spyros *et al.* The M5 uncertainty competition: Results, findings and conclusions. **International Journal of Forecasting**, 2021.

MITCHELL, Tom. **Machine Learning**. New York: McGraw-hill, 1997.

NAVIN, G. V. Big Data Analytics for Gold Price Forecasting Based on Decision Tree Algorithm and Support Vector Regression (SVR). **International Journal of Science and Research**, v. 4, n. 3, p. 2026–2030, 2015.

NIELSEN, Aileen. **Practical Time Series Analysis: Prediction with Statistics and Machine Learning**. California: O'Reilly Media, Inc, 2019.

OLIVEIRA, E. M., OLIVEIRA, F. L. Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods. **Energy**, v. 144, p. 776–788, 2018.

ORESHKIN, Boris *et al.* N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. In: **Eighth International Conference on Learning Representations**, Addis Ababa: 2020. Disponível em < <https://arxiv.org/abs/1905.10437> > .

ORESHKIN, Boris *et al.* N-BEATS neural network for mid-term electricity load forecasting. **Applied Energy**, v. 293, 2021.

PAI, Ping-Feng, HONG, Wei-Chiang. Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms. **Electric Power Systems Research**, v. 74, p. 417–425, 2005.

PHILLIPS, P.; PERRON, P. Testing for a unit root in autoregressive-moving average models of unknown order. **Biometrika**, v. 71, n. 3, p. 599–607, 1984.

RAFI, Shafiul *et al.* A Short-Term Load Forecasting Method Using Integrated CNN and LSTM Network. **IEEE Access**, v. 9, p. 32436–32448, 2021.

ROMERA, E. *et al.* Monthly Electric Energy Demand Forecasting with Neural Networks. **Energy Conversion and Management**, v. 49, p. 3135–3142, 2008.

SALINAS, David *et al.* DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks. **International Journal of Forecasting**, v. 16, n.3, p. 1181–1191, 2020.

SHIH, Shun-Yao *et al.* Temporal Pattern Attention for Multivariate Time Series Forecasting. **Machine Learning**, v. 108, n. 8, p. 1421–1441, 2019.

SHUMWAY, Robert, STOFFER, David. **Time Series Analysis and Its Applications: With R Examples**. 4th ed. New York: Springer, 2016.

SMYL, Slawek. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. **International Journal of Forecasting**, v. 36, n. 1, p. 75–85, 2020.

TAYLOR, Sean J.; LETHAM, Benjamin. Forecasting at scale. **The American Statistician**, v. 72, n. 1, p. 37–45, 2018.

UNSIHUAY-VILA, C. *et al.* Electricity demand and spot price forecasting using evolutionary computation combined with chaotic nonlinear dynamic model. **International Journal of Electrical Power & Energy Systems**, v. 32, n. 2, p. 108–116, 2010.

VELASQUEZ, Carlos *et al.* Analysis of time series models for Brazilian electricity demand forecasting. **Energy**, v. 247, 2022.

VEERAMSETTY, Venkataramana *et al.* Short Term Electric Power Load Forecasting Using Principal Component Analysis and Recurrent Neural Networks. **Forecasting**, v. 4, p. 149–164, 2022.

WADE, Corey. **Hands-On Gradient Boosting with XGBoost and scikit-learn**. Birmingham: Packt Publishing, 2020.

WICHARD, Jorg, OGORZALEK, Maciej. Time series prediction with ensemble models applied to the CATS benchmark. **IEEE International Joint Conference on Neural Networks**, v. 2, p. 1625-1630, 2004.

ZHENG, Zhuang *et al.* A Kalman filter-based bottom-up approach for household short-term load forecast. **Applied Energy**, v. 250, p. 882–894, 2019.

APÊNDICE A - ANÁLISE DA SÉRIE DE DEMANDA ELÉTRICA

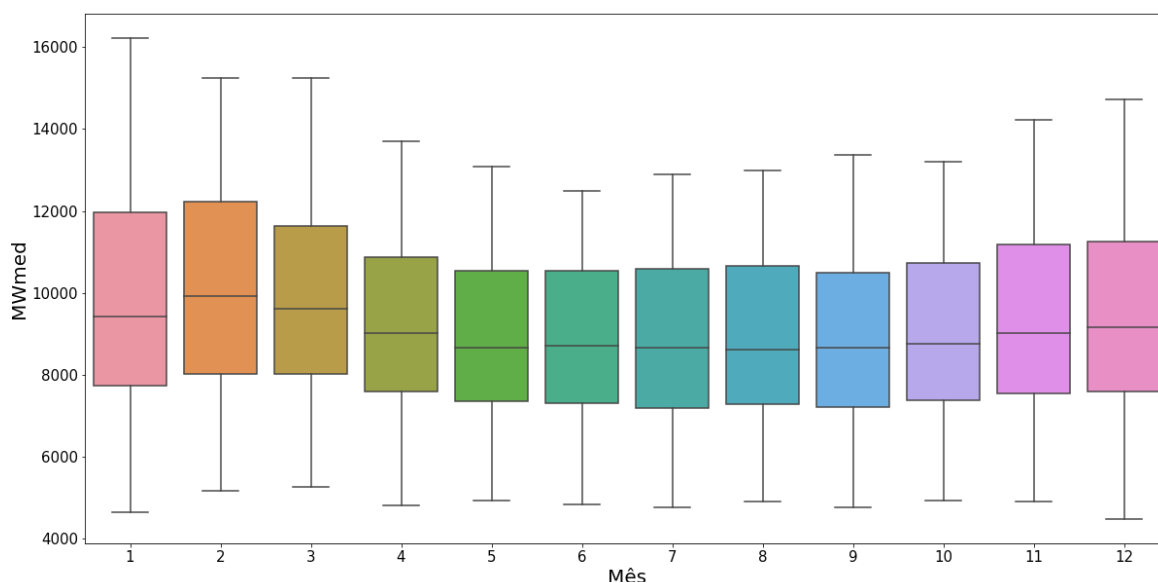


Figura 10 - Sazonalidade anual da demanda elétrica na região sul do Brasil.

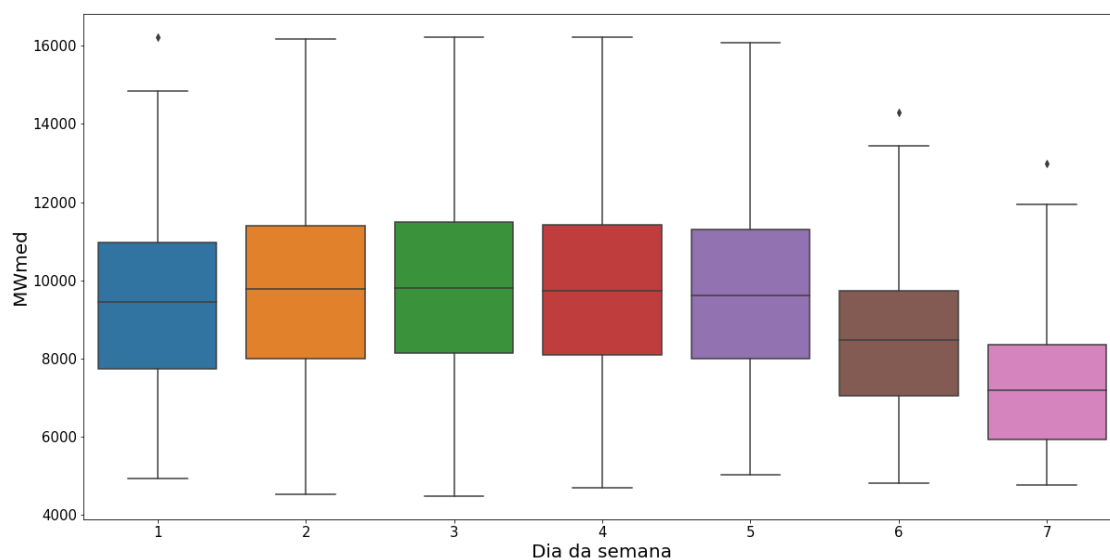


Figura 11 - Sazonalidade semanal⁷ da demanda elétrica na região sul do Brasil.

⁷ O número 1 corresponde à segunda-feira; o 7, ao domingo.

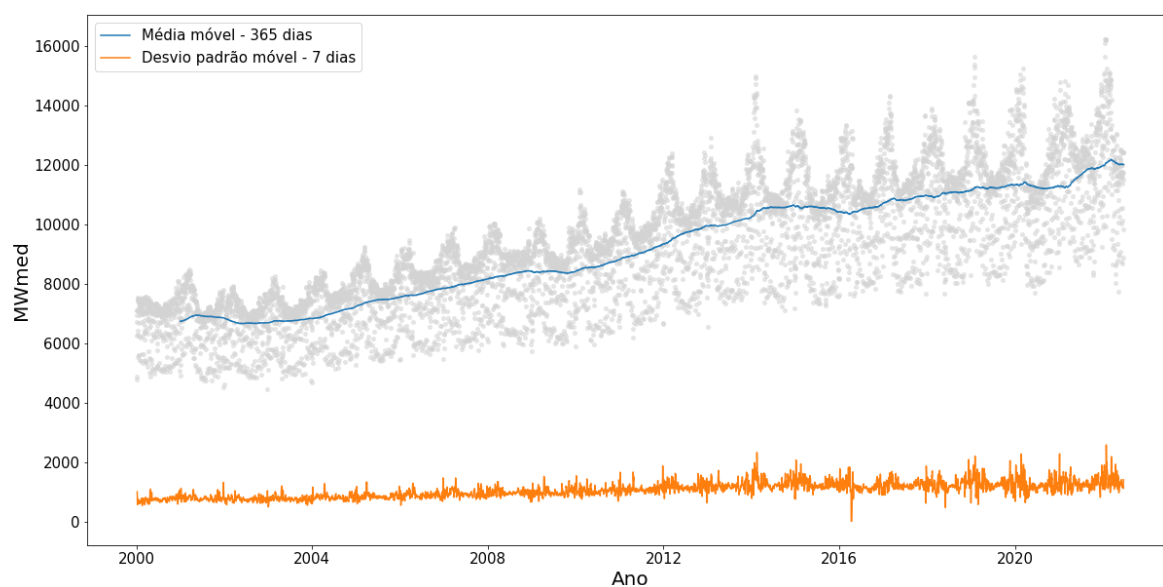


Figura 12 - Evolução da demanda elétrica na região sul do Brasil.

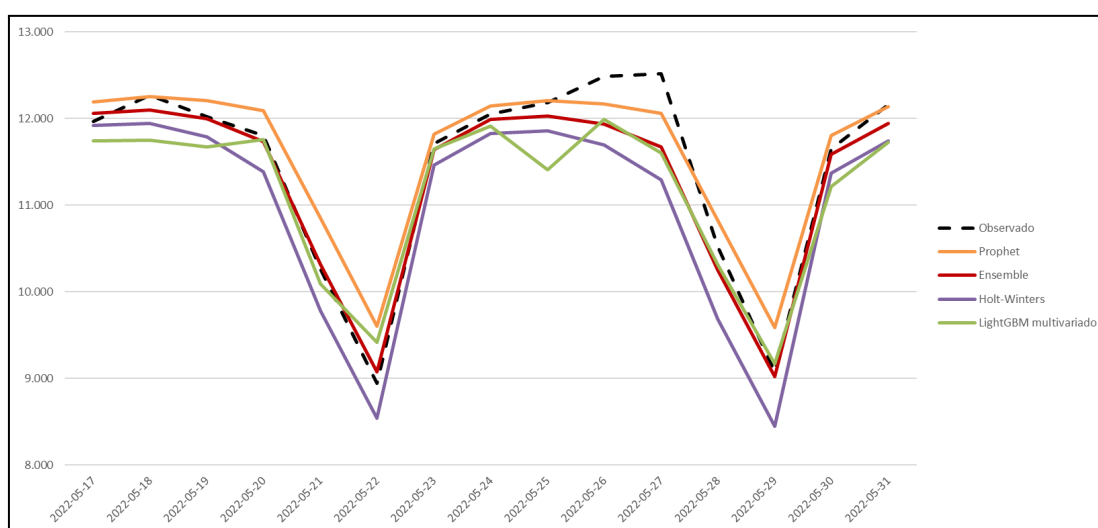


Figura 13 - Projeção dos quatro melhores modelos.