

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO RIO GRANDE
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL

O USO DOS MODELOS PARIMA E SARIMA NA PREVISÃO DA DEMANDA
MENSAL DE ENERGIA ELÉTRICA PARA A CLASSE RESIDENCIAL DO ESTADO
DO ESPÍRITO SANTO

por

Wara Martins Linhares

Dissertação para obtenção do Título de
Mestre em Modelagem Computacional

Rio Grande, setembro de 2023.

Ficha Catalográfica

L755u Linhares, Wara Martins.

O uso dos modelos PARIMA e SARIMA na previsão da demanda mensal de energia elétrica para a classe residencial do estado do Espírito Santo / Wara Martins Linhares. – 2023.
90 f.

Dissertação (mestrado) – Universidade Federal do Rio Grande – FURG, Programa de Pós-Graduação em Modelagem Computacional, Rio Grande/RS, 2023.
Orientadora: Dra. Viviane Leite Dias de Mattos.

1. Séries Temporais 2. Previsão 3. SARIMA 4. PARIMA
5. Energia elétrica I. Mattos, Viviane Leite Dias de II. Título.

CDU 004

Catálogo na Fonte: Bibliotecário José Paulo dos Santos CRB 10/2344

Wara Martins Linhares

“O USO DOS MODELOS PARIMA E SARIMA NA PREVISÃO DA DEMANDA MENSAL DE ENERGIA ELÉTRICA PARA A CLASSE RESIDENCIAL DO ESTADO DO ESPÍRITO SANTO”

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal do Rio Grande - FURG, como requisito parcial para obtenção do Grau de Mestra. Área de concentração: Modelagem Computacional.

Aprovada em: 19/09/2023.

BANCA EXAMINADORA

Documento assinado digitalmente
 **VIVIANE LEITE DIAS DE MATTOS**
Data: 20/10/2023 07:06:24-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. Viviane Leite Dias de Mattos (Orientadora – FURG)

Documento assinado digitalmente
 **ANDRE ANDRADE LONGARAY**
Data: 20/10/2023 08:17:08-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. André Andrade Longaray (FURG)

Documento assinado digitalmente
 **ANDREA CRISTINA KONRATH**
Data: 20/10/2023 08:10:22-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. Andréa Cristina Konrath (UFSC)

Documento assinado digitalmente
 **Luiz Ricardo Nakamura**
Data: 20/10/2023 08:35:43-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Luiz Ricardo Nakamura (UFLA)

Rio Grande RS
2023

Dedico esse trabalho a minha amada avó Germina, a minha mãe, Maria, e aos meus irmãos, com todo meu amor e gratidão, por tudo que fizeram por mim ao longo de minha vida. Desejo poder ter sido merecedora do esforço dedicado por vocês em todos os aspectos, especialmente quanto à minha formação.

AGRADECIMENTOS

Agradeço à coordenação de aperfeiçoamento pessoal de nível superior (CAPES) e aos órgãos financiadores CNPq e CAPES pelo incentivo financeiro que permitiu a realização de uma formação acadêmica de qualidade.

Agradeço a Deus, pela vida, pela saúde e por permitir a conclusão de mais uma etapa importante da minha vida.

A minha orientadora, Viviane Leite Dias de Mattos, por toda dedicação, paciência, incentivo e confiança durante o desenvolvimento deste trabalho.

Aos professores do Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal do Rio Grande (FURG).

RESUMO

Muitos estudos têm mostrado a importância do planejamento e previsão da demanda de energia elétrica, cujo objetivo é manter o equilíbrio entre a produção e o consumo desse produto. Entender a sazonalidade e as características individuais das subséries nos estudos de previsão é importante para se aproximar dos padrões e comportamentos dessas variáveis e reduzir os erros de previsão que impactam diretamente nas questões técnicas e financeiras. O objetivo deste estudo é usar o método de Box-Jenkins, Modelo Autorregressivo Integrado de Médias Móveis Sazonal (SARIMA) e Modelo Autorregressivo Integrado de Médias Móveis Periódico (PARIMA), para modelar a demanda de energia da classe residencial no Espírito Santo e identificar o melhor protocolo para a modelagem. Neste estudo foram utilizados dados históricos de janeiro de 2004 a dezembro de 2021. Inicialmente, foi realizada análise exploratória da série e subséries, sumarização dos dados por meio de técnicas analíticas e gráficas e análise de variância paramétrica com blocos (teste F) e não paramétrica (teste de Friedman), avaliando-se a presença de sazonalidade, o que levou à seleção dos modelos SARIMA e PARIMA. Testes de Dickey-Fuller aumentado (ADF), Kwiatkowski-Phillips-Schmidt-Shin (KPSS) e Phillips- Perron (PP) foram usados para avaliar a estacionaridade da parte não sazonal e os testes Hilleberg-Engle- Granger-Yoo (HEGY), Osborn- Chui- Smith- Birchenhall (OCSB) e Canova-Hansen (CH) para a parte sazonal. Os coeficientes dos modelos candidatos foram então estimados e os modelos mais parcimoniosos foram determinados usando os critérios de informação Bayesiano (BIC), Akaike (AIC) e Akaike corrigido (AICc). Na análise dos resíduos foram utilizados os testes de Ljung-Box, Jarque-Bera e ARCH, além dos indicadores: Erro Absoluto Médio (MAE), Erro Médio Percentual Absoluto (MAPE) e Raiz do Erro Quadrático Médio (RMSE) para avaliar a capacidade preditiva, complementados por intervalos de confiança. Dentre as técnicas utilizadas nas fases do ciclo iterativo composta pela análise exploratória, identificação, estimação, diagnóstico e validação, as que alcançam melhor desempenho são: gráficos de linhas da série e subséries, teste KPSS, determinação de valores mínimos e máximos para p e q, critérios AIC e AICc e RMSE com análise da distribuição residual. No geral, ambos os métodos apresentam bons resultados, mas de acordo com o indicador de qualidade RMSE e a análise de resíduos, o método SARIMA apresenta melhor capacidade preditiva, podendo contribuir para planejamento da matriz energética do estado do Espírito Santo.

Palavras-chave: Séries Temporais, Previsão, SARIMA, PARIMA, Energia elétrica.

ABSTRACT

Many studies have shown the importance of planning and forecasting electricity demand, whose objective is to maintain the balance between production and consumption of this product. Understanding seasonality and the individual characteristics of subspecies in forecasting studies is important to approximate the patterns and behaviors of these variables and reduce forecasting errors that directly impact technical and financial issues. The aim of this study is to use the Box-Jenkins method, Seasonal Autoregressive Integrated Moving Average (SARIMA) model, and Periodic Autoregressive Integrated Moving Average (PARIMA) model to model residential energy demand in the state of Espírito Santo and identify the best modeling protocol. Historical data from January 2004 to December 2021 were used in this study. Initially, an exploratory analysis of the series and subspecies from January to December was conducted, data were summarized using analytical and graphical techniques, and parametric analysis of variance (ANOVA with blocks) and non-parametric analysis (Friedman test) were performed to assess the presence of seasonality. Its led to the selection of the SARIMA and PARIMA models. Augmented Dickey-Fuller (ADF), Kwiatkowski-Phillips-Schmidt-Shin (KPSS), and Phillips-Perron (PP) tests were used to assess the stationarity of the non-seasonal part, and Hilleberg-Engle-Granger-Yoo (HEGY), Osborn-Chui-Smith-Birchenhall (OCSB), and Canova-Hansen (CH) tests were used for the seasonal part. The coefficients of candidate models were then estimated, and the most parsimonious models were determined using Bayesian information criteria, Akaike criteria, and corrected Akaike criteria. Residual analysis included Ljung-Box, Jarque-Bera, and ARCH tests, in addition to the following indicators: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) to assess predictive capacity, complemented by confidence intervals. Among the techniques used in the phases of the iterative cycle consisting of exploratory analysis, identification, estimation, diagnosis, and validation, the ones that perform best are line charts of the series and subspecies, KPSS test, determination of minimum and maximum values for p and q, AIC and AICc criteria, and RMSE with residual distribution analysis, respectively. Overall, both methods yield good results, but according to the RMSE quality indicator and residual analysis, the SARIMA method demonstrates better predictive capacity, which can contribute to the planning of the energy matrix in the state of Espírito Santo.

Keywords: Time series, Forecasting, SARIMA, PARIMA, Electricity

LISTA DE FIGURAS

Figura 2.1- Fluxograma para a realização da modelagem Box-Jenkins	23
Figura 3.1- Fluxograma das etapas metodológicas para modelagem PARIMA.....	42
Figura 3.2- Fluxograma das etapas metodológicas para modelagem SARIMA.....	43
Figura 4.1- Gráfico em linhas para a demanda mensal de energia elétrica para a classe residencial/ ES, 2004-21.....	44
Figura 4.2- Histograma e gráfico de caixa para a demanda mensal de energia elétrica para a classe residencial/ ES, 2004-21	45
Figura 4.3- Gráfico de caixa para cada ano da demanda mensal de energia elétrica para a classe residencial/ ES, 2004- 21	45
Figura 4.4- Gráfico em linhas para cada subsérie da demanda de energia elétrica residencial para os meses, ES/2004-21	48
Figura 4.5- Gráfico de caixa para cada subsérie da demanda mensal de energia elétrica para a classe residencial/ ES, 2004- 21	49
Figura 4.6- Gráfico da FAC e FACP para o mês de janeiro com grau de integração 1	53
Figura 4.7- Gráfico da FAC e FACP para o mês de fevereiro com grau de integração 1	53
Figura 4.8- Gráfico da FAC e FACP para o mês de março com grau de integração 1	53
Figura 4.9- Gráfico da FAC e FACP para o mês de abril com grau de integração 1	54
Figura 4.10- Gráfico da FAC e FACP para o mês de maio com grau de integração 1	54
Figura 4.11- Gráfico da FAC e FACP para o mês de junho com grau de integração 1	54
Figura 4.12- Gráfico da FAC e FACP para o mês de julho com grau de integração 1	55
Figura 4.13- Gráfico da FAC e FACP para o mês de agosto com grau de integração 2	55
Figura 4.14- Gráfico da FAC e FACP para o mês de setembro com grau de integração 2	55
Figura 4.15- Gráfico da FAC e FACP para o mês de outubro com grau de integração 1	56
Figura 4.16- Gráfico da FAC e FACP para o mês de novembro com grau de integração 1	56
Figura 4. 17- Gráfico da FAC e FACP para o mês de dezembro com grau de integração 1 ...	56
Figura 4.18- Gráfico do modelo de previsão PARIMA para demanda mensal de energia elétrica da classe residencial/ES, 2004 – 21	67
Figura 4.19- Gráficos da FAC e FACP com grau de integração 1, para a parte não sazonal..	71
Figura 4.20- Gráficos da FAC e FACP da série em nível para a parte sazonal.....	72
Figura 4.21- Histograma (a) e Box Plot (b) do modelo SARIMA(3,1,2)(1,0,3) ₁₂ para demanda mensal de energia elétrica da classe residencial/ES, 2021	77

Figura 4. 22- Histograma (a) e Box Plot (b) do modelo SARIMA(2,1,1)(1,0,3) ₁₂ para demanda mensal de energia elétrica da classe residencial/ES, 2021	77
Figura 4.23- Gráfico do modelo de previsão SARIMA(2,1,1)(1,0,3) ₁₂ para demanda mensal de energia elétrica da classe residencial/ES, 2004 - 21.....	78
Figura 4.24- Gráfico de previsão do modelo SARIMA(0,1,1)(0,1,1) ₁₂ para a demanda mensal de energia elétrica da classe residencial, ES/ 2004-21	80
Figura 4.25- Gráfico da previsão para a demanda residencial ES/ 2021 gerado pelo modelo PARIMA	81
Figura 4.26- Gráfico da previsão para a demanda residencial ES/ 2021 gerado pelo modelo SARIMA	81

LISTA DE TABELAS

Tabela 4.1- Medidas resumo para a demanda mensal de energia elétrica para a classe residencial/ ES, 2004-21	46
Tabela 4.2- Resultado da ANOVA em blocos para a demanda mensal de energia elétrica para a classe residencial/ ES, 2004-21	46
Tabela 4.3- Análise dos resíduos ANOVA da demanda mensal de energia elétrica para a classe residencial/ ES, 2004- 21	47
Tabela 4. 4- Resultado do teste de Friedman para a demanda mensal de energia elétrica para a classe residencial/ ES, 2004-21	47
Tabela 4.5- Medidas resumo para as doze subséries da demanda mensal de energia elétrica para a classe residencial/ ES, 2004-21	50
Tabela 4.6- Resultados dos testes de estacionariedade para cada subsérie da demanda mensal de energia elétrica/ES, 2004 – 21.....	52
Tabela 4.7- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de janeiro, ES, 2004-2021.....	57
Tabela 4.8- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de fevereiro, ES, 2004-2021	57
Tabela 4.9- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de março, ES, 2004-2021.....	58
Tabela 4.10- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de abril, ES, 2004-2021	58
Tabela 4.11- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de maio, ES, 2004-2021.....	58
Tabela 4.12- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de junho, ES, 2004-2021.....	59
Tabela 4.13- Resultados dos critérios de informação e da análise de resíduos dos modelos	

candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de julho, ES, 2004-2021	59
Tabela 4.14- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de agosto, ES, 2004-2021	60
Tabela 4. 15- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de setembro, ES, 2004-2021	60
Tabela 4.16- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de outubro, ES, 2004-2021	61
Tabela 4. 17- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de novembro, ES, 2004-2021	61
Tabela 4.18- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de dezembro, ES, 2004-2021	61
Tabela 4.19- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de janeiro, ES, 2021	62
Tabela 4.20- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de fevereiro, ES, 2021	62
Tabela 4.21- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de março, ES, 2021.....	63
Tabela 4.22- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de abril, ES, 2021	63
Tabela 4.23- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de maio, ES, 2021.....	63
Tabela 4.24- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de junho, ES, 2021	63
Tabela 4.25- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de julho, ES, 2021	64
Tabela 4.26- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de agosto, ES, 2021	64
Tabela 4.27- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de	

energia elétrica da classe residencial no mês de setembro, ES, 2021	64
Tabela 4.28- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de outubro, ES, 2021	65
Tabela 4.29- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de novembro, ES, 2021	65
Tabela 4.30- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de dezembro, ES, 2021	65
Tabela 4.31- Capacidade preditiva do Modelo PARIMA da demanda mensal de energia elétrica para a classe residencial e respectivas medidas de acurácia, ES/2021.....	66
Tabela 4.32- Valores observados da demanda mensal de energia elétrica para a classe residencial e valores previstos e intervalo de confiança obtidos a partir de modelo PARIMA, ES, 2021	68
Tabela 4.33- Capacidade preditiva do Modelo PARIMA da demanda mensal de energia elétrica para a classe residencial, gerado a partir da função auto.arima, e respectivas medidas de acurácia, ES/2021.....	69
Tabela 4.34- Valores observados da demanda mensal de energia elétrica para a classe residencial e valores previstos e intervalo de confiança obtidos a partir de modelo PARIMA gerado a partir da função auto.arima, ES, 2021	70
Tabela 4.35- Resultado do teste de estacionariedade da demanda mensal de energia elétrica para a classe consumidora residencial/ES, 2004 – 21	71
Tabela 4.36- Resultado do teste de estacionariedade da demanda mensal de energia elétrica para a classe consumidora residencial/ES, 2004 – 21	72
Tabela 4.37- Modelos candidatos escolhidos pela parcimônia para a demanda mensal de energia elétrica para a classe consumidora residencial, ES/2004-21	73
Tabela 4.38- Indicadores de qualidade para validação dos modelos SARIMA da classe consumidora residencial/ ES,2004-21.....	74
Tabela 4.39- Valores observados da demanda mensal de energia elétrica para a classe residencial e valores previstos e intervalo de confiança obtidos a partir de modelo SARIMA(3,1,2)(1,0,3) ₁₂ , ES,2021	75
Tabela 4.40- Valores observados da demanda mensal de energia elétrica para a classe residencial e valores previstos e intervalo de confiança obtidos a partir de modelo SARIMA(2,1,1)(1,0,3) ₁₂ , ES,2021	76
Tabela 4.41- Resultados dos critérios de informação AIC, AICc e BIC para os modelos considerados como melhores pela função auto.arima para a demanda de energia elétrica da	

classe residencial/ ES,2004-21.....	78
Tabela 4.42- Valores observados da demanda mensal de energia elétrica para a classe residencial e valores previstos e intervalo de confiança obtidos a partir de modelo SARIMA(0,1,1)(0,1,1) ₁₂ , ES, 2021	79
Tabela 4.43- Capacidade preditiva obtidas nas modelagens PARIMA e SARIMA.....	80

LISTA DE SÍMBOLOS

A	Coeficiente de assimetria
B	Operador de retardo
c	Constante
C	Coeficiente de curtose
d	Número de diferenças
D	Número de diferenças sazonais
$E[.]$	Esperança matemática
h	Quantidade de erros passados
K	Defasagens
LM	Estatística do teste KPSS
$\text{Log}(L)$	Logaritmo da função de verossimilhança
m	Número máximo de defasagens utilizadas
N	Tamanho da amostra analisada
p	Ordem do processo autorregressivo ($i = 1, 2, \dots, p$)
P	Ordem do processo autorregressivo sazonal
q	Ordem do processo de médias móveis ($j = 1, 2, \dots, q$)
Q	Ordem do processo de médias móveis sazonal
QM	Quadrado médio
s	Número de períodos dentro do ano
S_t	Soma parcial dos erros
SQ	Somas de quadrados
SQE	Soma do quadrado dos erros
W	Vetor de parâmetros populacionais da função de máxima verossimilhança
Y_t	t-ésimo valor observado ($t = 1, 2, \dots, N$)
$\bar{Y}$	Média das observações

$Y_{v,\tau}$	Variável dependente para o ano v e o período τ , com $\tau = 1, \dots, s$
$\hat{\sigma}_{\varepsilon}^2$	Estimativa da variância ao longo prazo de Y_t
μ	Média das observações
μ_{τ}	Média do τ -ésimo período
β_1	Coeficiente associado a constante
α_i	Estatística de teste utilizada no teste DF
β_2	Coeficiente de tendência
δ	Medida da raiz unitária no teste DF
ε_t	Erro no t -ésimo tempo
θ_j	Coeficiente de médias móveis na j -ésima defasagem
$\theta_{j,\tau}$	Coeficientes periódicos de média móvel
ρ_i	Coeficiente de autocorrelação para um modelo AR(p)
ρ_j	Coeficiente de autocorrelação para um modelo MA(q)
φ_i	Coeficiente da parcela auto-regressiva na i -ésima defasagem
$\varphi_{i,\tau}$	Coeficientes periódicos autorregressivos
∇^n	n -ésima diferença de Y
γ_k	Autocovariância com k defasagens
τ	Estatística de teste utilizada no teste DF
$\hat{\sigma}_{\varepsilon}^2$	Estimativa da variância ao longo prazo de Y_t
π	Coeficiente utilizados no teste de raiz unitária HEGY
$y_{i,t}$	Resultado do trimestre
ψ_1, ω_1	Coeficiente associados a diferenças de ordem S da parte não sazonal
ψ_2, ω_2	Coeficientes associados a diferenças de ordem 1 da parte sazonal
$r_{t,t-k}$	Coeficientes de autocorrelação entre as observações dos períodos com k defasagens

LISTA DE ABREVIATURAS

ABSOLAR	Associação brasileira de energia fotovoltaica
ADF	Dickey-Fuller aumentado
AIC	Critério de informação de Akaike
AICc	Critério de informação de Akaike corrigido
ALM	Teste ARCH multiplicador de Lagrange
ANEEL	Agência Nacional de Energia Elétrica
ARIMA	Autorregressivo integrado e de média móvel
ARMA	Autorregressivo com média móvel
BEES	Balanco energético do estado do Espírito Santo
BIC	Critério de informação Bayesiano
EPE	Empresa de Pesquisa Energética
FAC	Função de autocorrelação
FACP	Função de autocorrelação parcial
IC	Intervalo de confiança
JB	Jarque-Bera
LB	Ljung-Box
MAE	Erro Absoluto Médio
MAPE	Erro Médio Percentual Absoluto
PARIMA	Autorregressivo integrado e de média móvel periódico
RMSE	Raiz do Erro Quadrático Médio
SARIMA	Autorregressivo integrado e de média móvel sazonal
VAR	Vetores autorregressivos
VAR-VEC	Vetores autorregressivo em conjunto com Correções de Erros Vetoriais

SUMÁRIO

1.INTRODUÇÃO.....	17
1.2.Justificativa.....	18
1.3. Objetivos	19
1.3.1. Objetivo geral.....	19
1.3.2. Objetivos específicos.....	20
1.4.Estrutura do trabalho	20
2. REFERENCIAL TEÓRICO	21
2.1.Modelos de Box- Jenkins	21
2.2.Procedimentos da modelagem Box-Jenkins	23
2.2.1. Testes de estacionariedade	24
2.2.2. Gráficos das funções de autocorrelação e autocorrelação parcial....	30
2.2.3. Estimacão e critérios de informacão	31
2.2.4. Análise dos resíduos.....	32
2.2.5. Capacidade preditiva	33
3. METODOLOGIA	40
3.1.Procedimentos metodológicos na análise exploratória.....	40
3.2.Procedimentos metodológicos na modelagem PARIMA.....	41
3.3.Procedimentos metodológicos na modelagem SARIMA.....	42
4. RESULTADOS E DISCURSÃO.....	44
4.1.Análise exploratória	44
4.2.Modelagem PARIMA	50
4.3.Modelagem SARIMA	71
4.4.Considerações sobre os resultados das modelagens PARIMA e SARIMA	80
4.5.Considerações sobre o desempenho das técnicas utilizadas nas modelagens PARIMA e SARIMA	82
5. CONSIDERAÇÕES FINAIS	84
REFERÊNCIAS	85

1. INTRODUÇÃO

1.1. Apresentação do tema.

À medida que os padrões de vida das pessoas melhoram, a demanda por eletricidade aumenta, e as quedas de energia em algumas áreas se tornarão um problema, a menos que sejam feitos planos que levem em consideração as necessidades futuras (YANG *et al.* 2016). Stern *et al.* (2019) correlacionaram o crescimento da riqueza ao longo do tempo com o aumento do uso de energia na forma de eletricidade nos países ricos, sugerindo que este aumento pode ser usado como um indicador de desenvolvimento. Um estudo da relação entre o consumo de eletricidade e o desenvolvimento econômico brasileiro, realizado com dados entre 1970 e 2000, mostrou que o crescimento do PIB esteve fortemente associado ao consumo de eletricidade neste período (LUZ, 2003). No geral, a relação entre o consumo de eletricidade e o desenvolvimento econômico mostra uma relação mais forte nos países mais ricos do que nos mais pobres (STERN *et al.* 2019).

É sabido que a eletricidade é um recurso difícil de armazenar e que a sua produção é afetada por diversos fatores, como: clima, população, emergências e até sazonalidade. Dadas as questões acima, estimar a carga necessária em um determinado momento é uma tarefa complexa, que requer previsões precisas e o uso de métodos de previsão. Os modelos de previsão são, portanto, ferramentas promissoras para prever a produção, demanda e reservas de energia (YANG *et al.* 2016).

O Brasil gera 85% da eletricidade de sua matriz energética a partir de fontes renováveis. Embora a capacidade energética das usinas hidrelétricas seja altamente dependente da variabilidade ambiental, a dependência da energia hidrelétrica é alta, pois esta é responsável por mais de 76% do fornecimento de eletricidade (MARQUES, 2014).

Marcado pela escassez de chuvas em grande parte do território brasileiro, o ano de 2021 apresentou dificuldade em manter a produção de energia elétrica, não só pela dependência do clima, mas também pela falta de planejamento estratégico no setor, tendo como consequência uma crise no abastecimento de energia elétrica (BORGES, 2021). Por outro lado, a urgência em complementar estas fontes de geração de eletricidade pela matriz energética brasileira, associado ao fato da energia solar ser uma fonte limpa e renovável de eletricidade, fez com que a energia proveniente da luz solar começasse a ganhar um maior destaque na matriz energética brasileira (SANTOS; JABBOUR, 2013).

Mediante esse cenário, o poder público de alguns estados brasileiros tem adotado algumas medidas para incentivar o uso de fontes renováveis diversificadas para a geração de

eletricidade. Como forma de incentivo ao uso de fontes renováveis de energia, o governador do estado do Espírito Santo, por exemplo, assinou um decreto de lei que institui a obrigatoriedade da instalação de equipamentos para a captação de energia solar em novas edificações estaduais, através de instalação de placas fotovoltaicas no local da edificação (PAGOTTO, 2019). Neste estado também foi criado um Projeto de Lei (PL) 587/2020 que instituiu o Programa de Geração de Energias Renováveis do Espírito Santo (GERAR), com objetivo promover a diversificação da matriz energética estadual, a descentralização e interiorização do desenvolvimento socioeconômico, estimulando atividades que utilizem fontes de energias renováveis, contribuindo para a preservação do meio ambiente. Além disso, este programa tem outros objetivos, como estimular pesquisas voltadas ao desenvolvimento tecnológico de fontes de energias renováveis e o desenvolvimento do mercado de energia solar fotovoltaica no Estado (FREITAS, 2021).

Segundo a Associação Brasileira de Energia Fotovoltaica (ABSOLAR), a geração de energia solar no estado do Espírito Santo mais que dobrou no ano de 2022, com sua capacidade produtiva em residências, estabelecimentos comerciais, indústrias, imóveis rurais e prédios públicos atingindo o valor 341,2 MW. Em janeiro de 2022, a capacidade produtiva instalada de energia solar distribuída atingiu 146,5 megawatts (MW), representando um aumento de 133% e colocando o estado em 16º lugar no ranking nacional, além de alcançar mais de 32 mil consumidores na região do Espírito Santo.

Portanto, um modelo de previsão do consumo de energia elétrica em uma região pode fornecer subsídios para a elaboração de um planejamento de curto, médio e longo prazo de geração, transmissão e distribuição de energia elétrica, que poderá beneficiar muito o setor de produção de energia renovável, acompanhando o crescimento econômico. Assim, os modelos de previsão tornam-se ferramentas importantes para a previsão de produção, demanda e reservas de energia (POLPRASERT; NGUYÊN; CHAROENSOK, 2021).

Partindo do princípio de que um modelo do consumo/demanda de energia elétrica pode auxiliar a gestão da matriz energética de uma localidade, seria possível modelar o consumo/demanda de energia elétrica do estado do Espírito Santo para fornecer subsídios ao GERAR? Modelos obtidos pela metodologia de Box-Jenkins poderiam ser úteis?

1.2. Justificativa

O consumo de eletricidade no setor residencial brasileiro aumentou significativamente, apesar dos esforços para promover a eficiência energética em eletroeletrônicos. (TEIXEIRA *et al.*, 2022). Como o crescimento da demanda de energia está

diretamente relacionado ao desenvolvimento regional e o setor residencial é o segundo maior consumidor, é prudente planejar a expansão dos recursos essenciais para geração, transmissão e distribuição de energia neste setor. (DE OLIVEIRA *et al.*, 2010).

Segundo Matos *et al.* (2021), a análise do consumo/demanda de energia elétrica é uma etapa importante no planejamento de sua produção, distribuição e transmissão, sendo essencial para o funcionamento de diversos aspectos da gestão. No Espírito Santo, segundo o Balanço Energético do Estado do Espírito Santo (BEES, 2022) – ano base 2021, a matriz Energética Capixaba apresentou aumento da demanda no setor residencial e aumento do uso de fontes renováveis de energia, especialmente, a energia solar (ABSOLAR, 2021). Para atender a esta crescente demanda energética, o governo do estado do Espírito Santo direciona esforços para estimular o desenvolvimento da cadeia produtiva e do mercado de energia renovável no estado por meio de projetos como o GERAR.

Dessa forma, se faz necessário modelos preditivos, pois eles podem fornecer subsídios para o desenvolvimento de planos de curto, médio e longo prazos, que podem trazer benefícios significativos para a produção de energia renovável. (RINGKJØB; HAUGAN; SOLBREKKE, 2018). Diante deste cenário, a modelagem da demanda de eletricidade não apenas caracteriza seus padrões comportamentais, mas também faz previsões que podem fornecer subsídios para projetos de infraestrutura regional de forma a não atrapalhar ou desacelerar o desenvolvimento econômico da região (NUNES *et al.*, 2022).

Embora vários estudos tenham sido desenvolvidos para modelar a demanda energética do Brasil, até o momento não foi encontrado na literatura científica pesquisada nenhum trabalho específico sobre os dados de demanda energética do Espírito Santo no setor residencial e a criação de modelos preditivos com base nesses dados justifica a importância deste estudo. Vale ressaltar que existe um intervalo de tempo esperado de três a cinco anos entre o reconhecimento da necessidade de aumento da geração e o consumo real, evidenciando a necessidade de um sistema de planejamento adequado (FERREIRA; BRAGA; MENDES, 2006).

1.3. Objetivos

1.3.1. Objetivo geral

O presente estudo tem como objetivo modelar a demanda mensal de energia elétrica para a classe residencial de consumidores no estado do Espírito Santo, utilizando as abordagens PARIMA e SARIMA oriundas da metodologia de Box- Jenkins.

1.3.2. Objetivos específicos

- Conhecer por meio de uma análise exploratória algumas propriedades dos dados que compõem a série temporal analisada.
- Modelar a referida série utilizando dois métodos oriundos da metodologia de Box-Jenkins: modelos autorregressivos integrados de médias móveis sazonais (SARIMA) e modelos autorregressivos integrados de médias móveis com periodicidade (PARIMA).
- Identificar e aplicar diferentes protocolos que possam ser utilizados na operacionalização das diversas técnicas empregadas nos métodos selecionados, identificando o de melhor desempenho nestes dados, quando possível.
- Encontrar o modelo mais parcimonioso e com melhor capacidade preditiva, descrevendo o protocolo utilizado na sua determinação.
- Refletir sobre algumas técnicas (vantagens e desvantagens) utilizadas nas modelagens.

1.4. Estrutura do trabalho

O trabalho está organizado em cinco capítulos. No primeiro capítulo é introduzido o tema, mostrando sua importância, a justificativa para elaboração deste estudo, os objetivos e a descrição da estrutura do trabalho. No segundo capítulo apresenta-se o referencial teórico com os principais conceitos utilizados para o desenvolvimento deste trabalho, além de alguns estudos relacionados. A metodologia utilizada no trabalho encontra-se no capítulo três. No capítulo quatro são apresentados os resultados e discussão dos modelos encontrados nas abordagens PARIMA e SARIMA para a classe consumidora residencial. E por fim, no capítulo cinco encontram-se as considerações finais sobre o estudo realizado.

2. REFERENCIAL TEÓRICO

Este capítulo apresenta os conceitos básicos mais importantes da metodologia de Box-Jenkins. Posteriormente são apresentados os procedimentos relacionados à modelagem, seguidos de uma apresentação e discussão de alguns estudos relacionados com este tema.

2.1. Modelos de Box- Jenkins

A metodologia de Box-Jenkins se baseia na construção de funções que usam dados da própria série para encontrar uma estrutura que permita fazer previsões (GUJARATI & PORTER, 2011). Dentre as diferentes abordagens desta metodologia, temos: ARMA (autoregressive-moving-average), ARIMA (autoregressive integrated moving average), SARIMA (seasonal ARIMA) e PARIMA (periodic ARIMA).

Um modelo autorregressivo de médias móveis de ordens p e q , denominado ARMA(p,q), conforme Morettin (2011), pode ser definido pela Equação 2.1:

$$Y_t - \mu = \sum_{i=1}^p \varphi_i (Y_{t-i} - \mu) + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t, \quad 2.1$$

em que: Y_t é o t -ésimo valor observado ($t = 1, 2, \dots, N$), μ é uma constante, p é a ordem do processo autorregressivo ($i = 1, 2, \dots, p$), φ_i é o coeficiente da parcela auto-regressiva na i -ésima defasagem, θ_j é o coeficiente de médias móveis na j -ésima defasagem, q é a ordem do processo de médias móveis ($j = 1, 2, \dots, q$) e ε_t é o erro no t -ésimo tempo.

Quando a série é dita não estacionária, se faz necessário aplicar diferenças sucessivas para a obtenção da estacionariedade, assim, um valor d é adicionado ao modelo ARMA para a remoção da raiz unitária, gerando então o modelo ARIMA(p,d,q).

O operador de retardo B , também chamado de backshift ou translação para o passado, retarda uma certa quantidade de períodos de uma série temporal. Assim temos para k defasagens, $B^k Y_t = Y_{t-k}$. (MORETTIN e TOLOI, 2018; CRYER e CHAN, 2008). A utilização desta simbologia simplifica os cálculos para estimação dos coeficientes do modelo.

A Equação 2.2 mostra a aplicação da primeira diferença nos dados de uma série temporal:

$$\nabla Y_t = (1 - B)Y_t = Y_t - Y_{t-1}. \quad 2.2$$

A n -ésima diferença de Y_t é obtida recursivamente pela Equação 2.3:

$$\nabla^n Y_t = (1 - B)^n Y_t \quad 2.3$$

Assim, o modelo ARMA(p,q) com diferenças sucessivas pode ser reescrito conforme a Equação 2.4:

$$(1 - \varphi_1 B - \dots - \varphi_p B^p)(1 - B)^d Y_t = c + (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t, \quad 2.4$$

em que: O termo $(1-B)$ se refere a ordem de integração para obter a condição de estacionariedade da série (HYNDMAN; ATHANASOPOULOS, 2018).

Em dados que apresentam sazonalidade, ou seja, informações atípicas que ocorrem em períodos específicos do ano, é necessário acrescentar ao modelo ARIMA termos sazonais. Gera-se assim o modelo ARIMA sazonal ou SARIMA. Portanto, um modelo SARIMA(p,d,q)(P,D,Q)_m contém uma parte não sazonal com parâmetros (p,d,q) e uma parte sazonal com parâmetros (P,D,Q), em que m representa o período sazonal. Logo, o modelo geral para a série com essa característica pode ser descrito conforme a Equação 2.5:

$$\varphi(B)\phi(B^S)(1-B)^d(1-B^S)^D y_t = \theta(B)\Theta(B^S) \varepsilon_t \quad 2.5$$

em que:

$\varphi(B) = 1 - \varphi_1(B) - \dots - \varphi_p B^p$ é o operador autorregressivo estacionário de ordem p;

$\theta(B) = 1 - \theta_1(B) - \dots - \theta_q B^q$ é o operador de médias móveis invertível de ordem q;

$\phi(B^S) = 1 - \phi_1(B^S) - \dots - \phi_p B^{Sp}$ é o operador autorregressivo sazonal de ordem p, estacionário;

$\Theta(B^S) = 1 - \Theta_1(B^S) - \dots - \Theta_Q B^{SQ}$ é o operador de médias móveis sazonal de ordem Q, invertível (HYNDMAN; ATHANASOPOULOS, 2018).

Algumas séries históricas apresentam em sua estrutura autocorrelações que não dependem apenas do intervalo de tempo entre as observações, mas também do período observado. Os modelos autorregressivos de médias móveis com p parâmetros periódicos, ou seja, ARMA periódico ou PARMA são aqueles em que um ou todos os parâmetros variam ao longo de um período, dessa forma, um modelo PARMA(p,q) é um modelo ARMA (p,q) para cada período de estudo, sendo descrito na Equação 2.6:

$$Y_{v,o} - \mu_o = \sum_{i=1}^p \varphi_{i,o} (Y_{v,o-i} - \mu_{o-i}) - \sum_{j=1}^q \theta_{j,o} \varepsilon_{v,o-j} + \varepsilon_{v,o}, \quad 2.6$$

no qual: $Y_{v,o}$ é a variável dependente para o ano v e o período o , com $o = 1, \dots, s$ sendo s o número de períodos dentro do ano, μ_o é a média do o -ésimo período, $\varphi_{i,o}$ são os coeficientes periódicos autorregressivos, $\theta_{j,o}$ são os coeficientes periódicos de média móvel, $\varepsilon_{v,o}$ é o erro. (MAIDMENT, 1993 apud COSER *et al.* 2009).

2.2. Procedimentos da modelagem Box-Jenkins

De acordo com Morettin e Toloi (2004), a metodologia de Box e Jenkins (1970) consiste em ajustar modelos autorregressivos integrados a médias móveis, ARIMA(p, d, q), a um conjunto de dados que constituirá a base para a construção do modelo mais adequado.

Na metodologia Box-Jenkins, a criação de um modelo ARIMA obedece a um ciclo iterativo, em que os componentes do modelo a princípio são escolhidos de acordo com os próprios dados, e posteriormente, sofrem ajustes até a obtenção de um modelo ótimo. Segundo Morettin e Toloi (2004), esse ciclo iterativo é composto por quatro etapas: identificação, estimação, verificação e predição, realizadas conforme fluxograma apresentado na Figura 2.1.

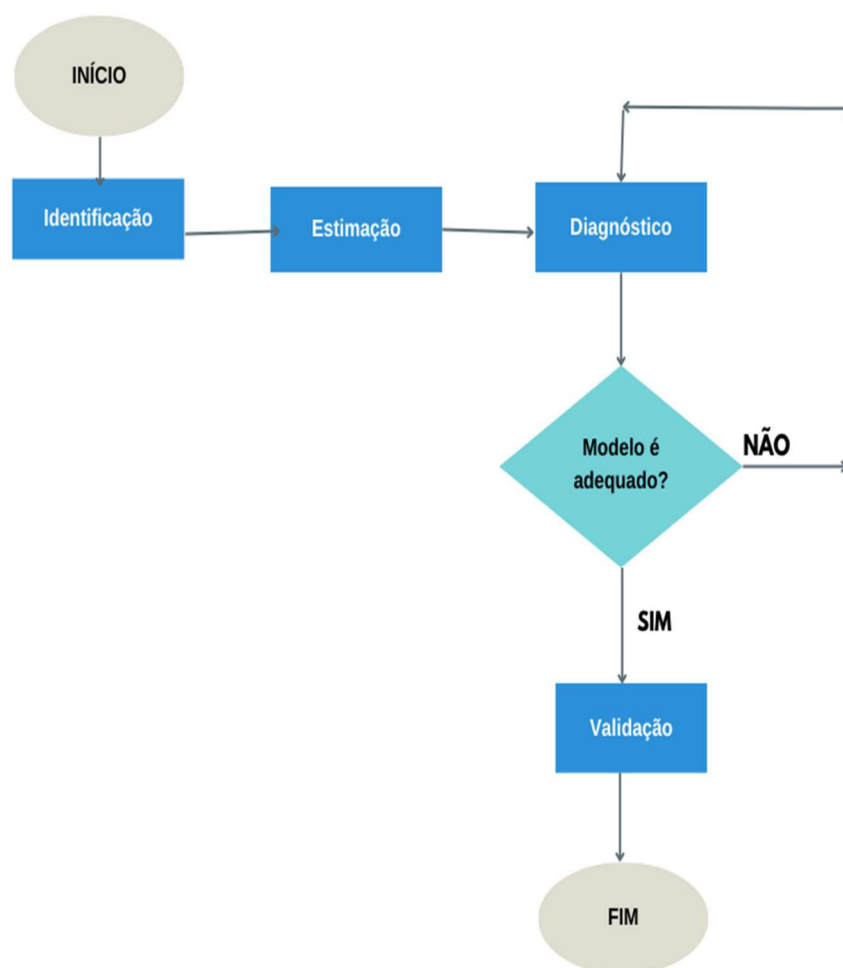


Figura 2.1- Fluxograma para a realização da modelagem Box-Jenkins

Fonte: Autora.

Antes da execução destas etapas, deve-se realizar uma análise exploratória, identificando algumas propriedades dos dados, assim como também avaliar a presença de raiz

unitária, tendo em vista que esta metodologia só deve ser aplicada em séries estacionárias.

As etapas identificadas por Morettin e Tolo (2004) são descritas a seguir.

Identificação: verifica qual modelo descreve o comportamento da série, utilizando a interpretação de correlogramas de funções de autocorrelação (FAC) e funções de autocorrelação parcial (FACP);

Estimação: estima dos parâmetros da componente autorregressiva e do componente de média móvel;

Diagnóstico: verifica se o modelo escolhido descreve apropriadamente o comportamento da série, o que é feito por meio da análise dos resíduos;

Validação: realização da previsão e determina a capacidade preditiva do modelo. Essa etapa representa o objetivo principal, realizada somente se as etapas anteriores forem satisfatórias. Assim, cada vez que o modelo não for adequado, as etapas devem ser repetidas até que se obtenha um modelo satisfatório que descreva o processo de geração da série (ZHANG, 2003; MORETTIN; TOLOI, 2004).

2.2.1. Testes de estacionariedade

Uma série temporal é dita estacionária se satisfaz as condições descritas nas equações 2.7, 2.8 e 2.9:

$$E(Y_t) = \mu, \forall t, \quad 2.7$$

$$E|Y^2| < \infty, \quad 2.8$$

$$cov(Y_t, Y_{t-k}) = \gamma_k, \quad 2.9$$

em que: Y_t é o valor observado e μ a média (MATTOS *et al.* 2021).

Segundo Morettin e Tolo (2018), os testes de raiz unitária ou testes de estacionariedade desempenham um importante papel na aplicação da modelagem de Box-Jenkins, pois além de identificar a presença de raiz unitária, determinam a ordem de integração (d) do modelo. Quando uma série apresenta raiz unitária, ela não é estacionária. Se ela se tornar ao tirar a primeira diferença, se diz que a mesma é integrada de ordem 1.

Existem muitos testes para avaliar a estacionariedade de uma série temporal, tais como: Dickey- Fuller (DF), Dickey-Fuller Aumentado (ADF), Kwiatkowski – Phillips – Schmidt – Shin (KPSS) e Phillips-Perron (PP) e, dentre estes, o mais usado na literatura pesquisada é o ADF. Silveira *et al.* (2017) sugerem que, em caso de não rejeição da hipótese nula no teste ADF, complementar a análise com o teste com o KPSS, pois possuem hipóteses nulas invertidas podendo ser considerados testes complementares.

- *Teste DF*

O teste de Dickey-Fuller ou teste DF usa o modelo apresentado na Equação 2.10, que é o ponto de partida de um processo estocástico.

$$Y_t = \varphi_1 Y_{t-1} + \varepsilon_t \quad \text{com } -1 \leq \varphi_1 \leq 1, \quad 2.10$$

em que: Y_t é a série a ser modelada, φ_1 é o coeficiente de Y_{t-1} e ε_t é o termo de ruído branco.

Quando $\varphi_1 = 1$, o processo estocástico é dito como não estacionário.

Para a realização do teste DF, a Eq. 2.10 pode ser trabalhada na forma de diferenças. Subtraindo Y_{t-1} em ambos os lados da igualdade, tem-se:

$$Y_t - Y_{t-1} = \varphi_1 Y_{t-1} - Y_{t-1} + \varepsilon_t, \quad 2.11$$

Logo:

$$Y_t - Y_{t-1} = (\varphi_1 - 1)Y_{t-1} + \varepsilon_t \quad 2.12$$

e pode ser reescrita conforme a Eq 2.13:

$$\nabla Y_t = \delta Y_{t-1} + \varepsilon_t \quad 2.13$$

em que: $\delta = \varphi_1 - 1$ e ∇ é o primeiro operador de diferença. (GUJARATI e PORTER, 2011).

Em Dickey e Fuller (1981) é demonstrado que, sob a hipótese nula de que $\delta = 0$, o valor estimado do coeficiente de Y_{t-1} na Eq. 2.9 não segue uma distribuição de t-student, todavia segue a distribuição de uma estatística denominada τ (tau) definida na Equação 2.14. Mediante a utilização da simulação de Monte Carlo, os autores computaram os valores fundamentais dessa estatística, definida conforme Eq. 2.14.

$$\tau = \frac{\delta}{EP(\delta)}, \quad 2.14$$

em que EP representa o erro padrão.

Na prática, testa-se a hipótese de que $\delta = 0$, o que acarreta $\varphi_1 = 1$, e por consequência, na presença de raiz unitária. Assim:

H_0 : $\delta = 0$ (o modelo possui raiz unitária, a série não é estacionária);

H_1 : $\delta < 0$ (o modelo não possui raiz unitária, a série é estacionária).

Chama-se a atenção para o fato de que um processo de passeio aleatório pode ser avaliado sob três perspectivas, considerando a inclusão de constante e tendência no modelo. Assim, temos três modelos a serem avaliados para o teste DF: considerando a inclusão de constante (β_1) e tendência (β_2), conforme Equação 2.15, considerando a inclusão de uma constante (β_1), conforme Equação 2.16, e considerando apenas o passeio aleatório, conforme Equação 2.17.

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \varepsilon_t \quad 2.15$$

$$\Delta Y_t = \beta_1 + \delta Y_{t-1} + \varepsilon_t \quad 2.16$$

$$\Delta Y_t = \delta Y_{t-1} + \varepsilon_t, \quad 2.17$$

- *Teste ADF*

Uma extensão deste teste é o de Dickey-Fuller Aumentado, sendo aplicado nos casos em que ε_t apresenta autocorrelação. Este teste parte do mesmo princípio do teste DF, porém considerando valores defasados da variável ΔY_t . Segundo Gujarati e Potter (2011), o teste ADF estima a regressão utilizando a Equação 2.18.

$$\nabla Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^m \alpha_i \nabla Y_{t-i} + \varepsilon_t, \quad 2.18$$

em que: β_1 é um termo independente chamado de constante (drift), β_2 é o coeficiente de tendência, δ é o coeficiente de presença de raiz unitária e m é o número máximo de defasagens utilizadas (GUJARATI e POTTER, 2011).

Esta quantidade máxima de defasagens (m_{max}) pode ser determinada conforme proposto em Schwert (1989), que faz uso da Equação 2.19, ou de acordo o proposto em Hyndman e Athanasopoulus (2018), que faz uso da Equação 2.20, ambas apresentadas a seguir:

$$m_{max} = \text{int} \left[12 * \left(\frac{N}{100} \right)^{\frac{1}{4}} \right] e \quad 2.19$$

$$m_{max} = \text{int} \left[4 * \left(\frac{N}{100} \right)^{\frac{2}{9}} \right], \quad 2.20$$

em que: N é o tamanho da série avaliada.

De acordo com Silveira (2019), a defasagem máxima encontrada pela Equação 2.20 conduz a melhores resultados, mesmo considerando menos valores passados. Salienta-se, entretanto, que o estudo desenvolvido por este autor não utilizou series com sazonalidade.

Para séries que apresentam sazonalidade, esse mesmo teste pode ser utilizado, dividindo-se a análise em duas etapas: inicialmente para a parte não sazonal com o objetivo de determinar a ordem de integração d , e posteriormente para a parte sazonal para determinar a ordem de integração sazonal D .

- *Teste KPSS*

O teste KPSS, proposto por Kwiatkowski *et al.* (1992), avalia a estacionariedade de

modo diferente ao teste ADF, em que:

$H0: \hat{\sigma}^2 = 0$ (o modelo não possui raiz unitária, a série é estacionária);

$H1: \hat{\sigma}^2 < 0$ (o modelo possui raiz unitária, a série é não estacionária).

A especificação do teste supõe definida a Equação 2.21, considerando um modelo com tendência, passeio aleatório e erro. De acordo com Silveira (2017), tem-se:

$$Y_t = \beta_2 t - r_t + \varepsilon_t, \quad 2.21$$

no qual: β_2 é uma constante que representa o coeficiente de t ; ε_t é um erro estacionário e um passeio aleatório; e $r_t = r_{t-1} + u_t$.

A estatística do teste é baseada na estatística LM (WANG, 2006), definida conforme Equação 2.22:

$$LM = \sum_{t=1}^T \frac{S_t^2}{N^2 \hat{\sigma}_\varepsilon^2}, \quad 2.22$$

em que: S_t é a soma parcial dos erros e $\hat{\sigma}_\varepsilon^2$ é a estimativa da variância ao longo prazo de Y_t , definidas pelas equações 2.23 e 2.24, respectivamente:

$$S_t = \sum_{i=1}^t e_i ; t = 1, 2, 3, \dots, N \quad 2.23$$

$$\hat{\sigma}_\varepsilon^2 = \frac{SQE}{N}, \quad 2.24$$

em que: SQE é a soma do quadrado dos erros.

- Teste PP

Apesar dos testes DF e ADF serem os mais utilizados, estes são conhecidos por terem um baixo poder estatístico, pois tem uma alta probabilidade de não rejeitar H_0 quando ela é falsa, ou seja, de assumir que as séries não são estacionárias quando na verdade o são (BUENO, 2011). Conforme Bueno (2011), o teste de Phillips-Perron (1988), chamado de teste PP ou teste z_{ϕ_1} , faz uma correção não paramétrica ao teste de Dickey e Fuller, que permite que este seja consistente mesmo quando há variáveis defasadas dependentes e os erros são correlacionados. O teste é baseado no seguinte modelo de regressão descrito pela Equação 2.25:

$$Y_t = \beta_1 + \phi_i Y_{t-1} + \varepsilon_t, \quad 2.25$$

e considera as hipóteses:

$H0: \phi_i = 1$ (o modelo possui raiz unitária, a série não é estacionária);

$H1: \phi_i < 1$ (o modelo não possui raiz unitária, a série é estacionária).

A estatística de teste é descrita na equação 2.26.

$$z_{\phi_1} = T(\hat{\phi}_1 - 1) - \frac{1}{2} \left(\frac{T^2 \hat{\sigma}_{\phi_1}^2}{S^2} \right) (\hat{\lambda} - \hat{\gamma}_0), \quad 2.26$$

em que: $\hat{\phi}_1$ é a estimação de ϕ_1 da Equação 2.26, $\hat{\sigma}_{\phi_1}^2$ é o erro padrão, $S^2 = (T - 2)^{-1} \sum_{t=1}^T \varepsilon_t^2$ em que ε_t é o resíduo da Equação 2.25 obtido por Mínimos Quadrados Ordinários (MQO), $\hat{\lambda}$ é uma estimativa da densidade espectral com frequência 0 de ε_t (variância de longo prazo de ε_t), $\hat{\gamma}_v$ é um estimador da covariância ($\hat{\gamma}_v = T^{-1} \sum_{v+1}^T (\hat{\varepsilon}_t \cdot \hat{\varepsilon}_{t-v})$).

- *Teste HEGY*

Para testar a presença de raízes unitárias em dados sazonais, um dos testes bastante utilizado é o teste HEGY, desenvolvido por Hylleberg *et al.* (1990). De acordo com Penna (2009), o procedimento HEGY avalia, em séries temporais lineares, a presença de raiz unitária isoladamente em todas as frequências sazonais, assim como na frequência zero. Ele testa a hipótese nula de que as raízes do polinômio autoregressivo pertencem ao círculo unitário, contra a alternativa de que estão fora. Este procedimento foi proposto inicialmente para avaliar a estacionariedade em dados trimestrais, sendo posteriormente adaptado para dados mensais.

Considere o operador diferença sazonal trimestral, apresentado na Equação 2.27, onde fica evidenciado a presença de quatro raízes unitárias (+1, -1, +i, -i).

$$(1 - B^4) = (1 - B)(1 + B)(1 - iB) = (1 - B)(1 + B)(1 + B^2) \quad 2.27$$

Assumindo que os dados são gerados por um modelo autoregressivo ($\varphi(B)Y_t = \varepsilon_t$), é possível demonstrar que, no caso de sazonalidade trimestral, tem-se:

$$\varphi(B)y_{4t} = \pi_1 y_{1,t-1} + \pi_2 y_{2,t-2} + \pi_3 y_{3,t-2} + \pi_4 y_{4,t-2} + \varepsilon_t \quad 2.28$$

em que:

$$y_{1,t} = (1 + B + B^2 + B^3)Y_t,$$

$$y_{2,t} = (1 - B + B^2 - B^3)Y_t,$$

$$y_{3,t} = (1 - B^2)Y_t,$$

$$y_{4,t} = (1 - B^4)Y_t.$$

Realizam-se então três testes para as raízes +1, -1 e $\pm i$, que consideram as seguintes hipóteses mostradas no Quadro 2.1

Quadro 2.1– Testes na realização do procedimento HEGY

Teste para raiz 1	Teste para raiz 2	Teste para raízes complexas
$H_0: \pi_1 = 0$	$H_0: \pi_2 = 0$	$H_0: \pi_3 = \pi_4 = 0$
$H_1: \pi_1 < 0$	$H_1: \pi_2 < 0$	$H_1: \pi_3, \pi_4 < 0$

Para concluir que uma série não tem raízes unitárias, sendo estacionária, Penna (2009) evidencia que é necessário que os π 's sejam diferentes de zero (entre π_3 e π_4 , pelo menos um). Mais detalhes deste teste podem ser encontrados em Hileberg *et al.* (1990) e Penna (2009).

-Teste OCSB

O teste de Osborn, Chui, Smith e Birchenhall [OCSB] (1988) foi desenvolvido a partir de Hasza e Fuller (1982), no contexto de teste para detectar presença diferenciada multiplicativa, ou seja, o teste OCSB investiga se ambos ou nenhum dos operadores $(1 - B)$ ou $(1 - E)$ devem ser aplicados aos dados. O modelo original é expresso na forma da equação 2.29, podendo ser generalizado com componentes determinísticos de acordo com a Equação 2.30. Logo:

$$\nabla_1 \nabla_s Y_t = \psi_1 \nabla_s Y_{t-1} + \psi_2 \nabla_1 Y_{t-s} + \varepsilon_t \quad 2.29$$

$$\eta(L) \nabla_1 \nabla_s Y_t = \mu_1 + \omega_1 \nabla_s Y_{t-1} + \omega_2 \nabla_1 Y_{t-s} + \varepsilon_t, \quad 2.30$$

em que: $\eta(L)$ é um polinômio AR (polinômio lag com raízes fora do círculo unitário),

$$\nabla_1 = (1 - B), \nabla_s = (1 - B^s) \text{ e } \mu_t = \beta_1 + \sum_{s=1}^{s-1} \alpha_s D_{s,t} + \beta_2 t + \sum_{s=1}^{s-1} \beta_s D_{s,t} t.$$

Para descobrir qual operador é adequado para y_t , as significâncias de ω_1 e ω_2 são testadas. Quando ambos ω_1 e ω_2 são considerados iguais a zero usando $\nabla_1 \nabla_s$, o modelo é adequado. Quando $\omega_1 = 0$ e $\omega_2 \neq 0$, o operador ∇_1 é selecionado. Quando $\omega_2 = 0$ e $\omega_1 \neq 0$, o operador ∇_s é selecionado. Se $\omega_1 \neq 0$ e $\omega_2 \neq 0$, nenhum operador diferencial é necessário (Franses, 1998: 563; Maddala and Kim, 1998: 366; Zhang, 2008: 11; Platon, 2010: 2-3).

-Teste CH

Canova e Hansen (1995), em um estudo apresentam os testes do multiplicador de Lagrange (LM). A estatística do teste CH é derivada do princípio LM que requer apenas a estimativa do modelo sob a hipótese nula, usando mínimos quadrados e são funções residuais

bastante simples. Esses testes também são utilizados para testar a estabilidade sazonal. O teste CH complementa os testes de Dickey et al.(1984) e Hylleberg et al. (1990) que investigaram as raízes unitárias sazonais com uma ou mais frequências sazonais. Assim, ao contrário desses testes de raiz unitária, hipótese nula do teste CH é que o processo é estacionário (ou seja, sazonalidade estacionária). A rejeição da hipótese nula implicaria na não estacionariedade dos dados (Canova e Hansen, 1995).

2.2.2. Gráficos das funções de autocorrelação e autocorrelação parcial

A autocorrelação ($r_{t,t-k}$), que nada mais é do que a autocovariância padronizada, mede a relação linear entre valores defasados da própria série, o que, de acordo com Hyndman e Athanasopoulus (2018), é feito conforme Equação 2.31:

$$r_{t,t-k} = \frac{\sum_{k+1}^N (Y_t - \bar{Y}) \cdot (Y_{t-k} - \bar{Y})}{\sum_1^N (Y_t - \bar{Y})^2}, \quad 2.31$$

em que: $r_{t,t-k}$ é a autocorrelação amostral, Y_t é a t -ésima observação ($t = 1, 2, \dots, N$), N é o tamanho da série, $\bar{Y}$ é a média das observações e k identifica a defasagem.

Esta medida representa a memória de um processo, ou seja, o quanto uma observação no tempo t depende da observação no tempo $t - i$.

Geralmente através da análise dos gráficos das funções de autocorrelação e autocorrelação parcial, construídos para uma série estacionária, realiza-se a etapa de identificação de um modelo ARIMA, na qual determinam-se as ordens dos termos autorregressivos (p) e das médias móveis (q) da série temporal.

De acordo com Thomaz (2019), a função de autocorrelação geral de um processo AR(p) é dada conforme a Equação 2.32, que deve decair exponencialmente.

$$\rho_i = \varphi_1 \rho_{i-1} + \varphi_2 \rho_{i-2} + \dots + \varphi_p \rho_{i-p}, \quad i = 1, 2, 3, \dots, p \quad 2.32$$

Na prática esta autocorrelação tende a desaparecer com o tempo. Em geral, quanto maior a distância temporal entre dois eventos, menor será a relação entre eles.

Para um modelo MA(q) a função de autocorrelação é definida conforme a Equação 2.33:

$$\rho_j = \frac{\theta_p + \theta_{j+1}\theta_1 + \dots + \theta_q\theta_{q-j}}{\sum_{j=0}^q \theta_j^2} \quad j=1, 2, 3, \dots, q \quad 2.33$$

$$\rho_j = 0, j > q$$

A FAC de um modelo MA puro é truncada na ordem q , pois se $j > q$, $\rho_j = 0$.

A interpretação das funções de autocorrelação e autocorrelação parcial, de acordo em Pereira (2020), pode ser utilizada para determinar as ordens p e q de um modelo ARIMA, conforme o Quadro 2.2:

Quadro 2.2- Resumo do comportamento gráfico das funções FAC e FACP

Modelo	FAC	FACP
AR(p)	Decai exponencialmente ou senoidalmente	Trunca na defasagem p
MA(q)	Trunca na defasagem q	Decai exponencialmente ou senoidalmente
ARMA(p,q)	Decai a partir de $p-q$	Decai a partir de $p-q$

2.2.3. Estimação e critérios de informação

Após a identificação das ordens da parte sazonal e não sazonal dos modelos candidatos, é feita a estimação de seus coeficientes, o que pode ser feito por diferentes abordagens. Segundo Moretin *et al.*, (2018), o Método da Máxima Verossimilhança é um dos mais utilizados em função de suas propriedades matemáticas. Os estimadores de máxima verossimilhança são os valores que maximizam a função de verossimilhança, sendo necessário supor que as inovações seguem uma distribuição de probabilidade. Se esta distribuição for Normal, com média zero e variância finita, as estimativas encontradas por este método coincidirão com aquelas encontradas pelo Método dos Mínimos Quadrados. Algumas vezes é necessário complementar o procedimento usando um método de otimização, como Nelder-Mead (NELDER; MEAD, 1965).

A escolha do melhor modelo entre os candidatos, do ponto de vista estatístico, é um item extremamente importante na análise de dados. Busca-se o modelo mais parcimonioso, isto é, o modelo que represente bem o comportamento da variável com a menor quantidade de coeficientes possíveis. O Critério de Informação de Akaike (AIC), o Critério de Informação de Akaike Corrigido (AICc) e o Critério de Informação Bayesiana (BIC) podem ser utilizados para evitar selecionar modelos com muitos parâmetros. São eles:

$$AIC = -2 \log(L) + 2(p + q) \quad 2.34$$

$$AICc = AIC + \frac{2(p + q)(p + q + 1)}{N - (p + q) - 1} \quad 2.35$$

$$BIC = -2 \log(L) + (p + q) \log(N), \quad 2.36$$

no qual: $\log(L)$ é o logaritmo da função de verossimilhança e N é o tamanho da amostra analisada. Em todos os critérios o modelo que apresentar menores valores é o modelo mais parcimonioso. Segundo Hyndman e Athanasopoulos (2018), em caso de divergência dos resultados, é preferível usar o AICc. Já Silveira (2023) aconselha a utilização do BIC.

2.2.4. Análise dos resíduos

Após estimar os parâmetros dos modelos candidatos, temos que verificar se eles representam, ou não, adequadamente, os dados. Tais verificações podem ser realizadas através de uma análise de resíduos, pois, a partir dos mesmos, teremos ideia se os dados satisfazem as suposições do modelo teórico, além de verificar se o modelo encontrado é válido. Os principais testes utilizados na verificação da independência, normalidade e homoscedasticidade dos resíduos são apresentados a seguir.

- *Teste Ljung-Box*

A primeira propriedade analisada nos resíduos é a presença de autocorrelação. Esta pode ser verificada pelo teste Ljung-Box (LB), proposto por Ljung e Box (1978), que testa se existe autocorrelação nos resíduos estimados.

As hipóteses do teste Ljung - Box são:

- H_0 : não existem evidências que os resíduos estejam autocorrelacionados, e,
- H_1 : existem evidências que os resíduos estejam correlacionados.

A estatística utilizada neste teste converge para uma distribuição χ^2 com $k-p-q$ graus de liberdade (CRYER e CHAN, 2008; BUENO, 2011; GUJARATI e PORTER, 2011) e pode ser encontrada conforme Equação 2.37:

$$LB = N(N + 2) \left(\frac{\hat{\rho}_1^2}{N-1} \right) + \left(\frac{\hat{\rho}_2^2}{N-2} \right) + \dots + \left(\frac{\hat{\rho}_k^2}{N-K} \right) \quad 2.37$$

em que: N é o tamanho da série, $\hat{\rho}_k$ é uma estimativa de autocorrelação com k defasagens ($k = 1, 2, \dots, K$).

- *Teste Jarque-Bera*

Proposto em Bera e Jarque (1980), o teste é utilizado para verificar a normalidade nos resíduos, uma propriedade desejável, e se baseia nos coeficientes de assimetria (A) e curtose (C) dos dados que compõe a série. As hipóteses do teste Jarque-Bera são:

- H_0 : não existem evidências que os resíduos não sigam uma distribuição Normal, e,
- H_1 : existem evidências que os resíduos não seguem uma distribuição Normal.

A estatística do teste é dada pela Equação 2.38:

$$JB = N \left| \frac{A^2}{6} + \frac{(C-3)^2}{24} \right|, \quad 2.38$$

esta estatística de teste também segue uma distribuição Quiquadrado (χ^2), mas com 2 graus de liberdade (CRYER e CHAN, 2008; BUENO, 2011; GUJARATI e PORTER, 2011).

-Teste ARCH

O teste ARCH (Autoregressive Conditional Heteroskedasticity) verifica a homocedasticidade dos resíduos do modelo. Ele é obtido a partir da Equação 2.39, que mostra um modelo de regressão linear de k variáveis, em que ε_t é normalmente distribuído com média zero e variância constante, conforme mostrado na Equação 2.40,

$$Y_t = \beta_0 + \beta_1 Y_{1t} + \dots + \beta_l Y_{lt} + \varepsilon_t \quad 2.39$$

$$\hat{\sigma}_\varepsilon^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2. \quad 2.40$$

Verifica-se na Equação 2.40, que a variância de ε no período t é dependente do quadrado do erro no tempo (t-1), sugerindo assim uma correlação serial. Além disso, a variância de ε_t pode depender de vários termos quadráticos defasados, como mostra a Equação 2.41,

$$\hat{\sigma}_\varepsilon^2 = \alpha_0 + \sum_{i=1}^h \alpha_i \varepsilon_{t-i}^2, \quad 2.41$$

em que: h é a quantidade de erros passados.

O teste ARCH considera que sua hipótese nula é que, na ausência de componentes ARCH (homocedasticidade), temos $\alpha_i = 0$ para todo $i = 1, \dots, h$, ou seja, todos os coeficientes da Equação 2.41 são nulos. A hipótese alternativa é que, na presença de componentes ARCH (heterocedasticidade), pelo menos um dos coeficientes α_i estimados deve ser diferente de zero.

Em uma amostra de N resíduos, por exemplo, a estatística de teste (Γ) é dada pela Equação 2.42,

$$\Gamma = N \cdot R^2, \quad 2.42$$

em que R^2 representa o coeficiente de determinação da regressão auxiliar obtida pela Equação 2.41. Esta estatística segue uma distribuição χ^2 com número de graus de liberdade igual ao número de coeficientes estimados na regressão auxiliar (GUJARATI e PORTER, 2011).

2.2.5. Capacidade preditiva

Segundo Scheidt *et al.* (2020), para avaliar a capacidade preditiva dos modelos, ou seja, determinar sua acurácia, utilizam-se usualmente as três medidas apresentadas a seguir.

- Erro Absoluto Médio (MAE – Mean Absolute Error)

O erro médio absoluto é calculado a partir da média dos erros absolutos, ou seja,

utilizamos o módulo de cada erro para evitar a subestimação, isso porque, o valor é menos afetado por outliers. Este indicador de qualidade mostra a distância entre os valores observados e previstos, apresentando a mesma unidade de medida da variável analisada.

$$MAE = \frac{1}{H} \sum_{i=1}^H |Y_t - \hat{Y}_t|, \quad 2.42$$

em que: Y_t e $\hat{Y}_t$ representam o valor real e o valor previsto, respectivamente, e H o número total de observações no período de validação da série.

- *Erro Médio Percentual Absoluto (MAPE – Mean Absolute Percentage Error)*

O MAPE, calculado conforme Equação 2.43, representa a média da diferença absoluta entre os valores previstos e observados, expressa em porcentagem dos valores observados. Este indicador de qualidade expressa em porcentagem a distância entre os valores observados e previstos.

$$MAPE = \left(\frac{1}{H} \sum_{t=1}^H \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \right) \quad 2.43$$

- *Raiz do Erro Quadrático Médio (RMSE – Root Mean Squared Error)*

RMSE é a medida que calcula "a raiz quadrática média" dos erros entre valores observados e valores previsto, conforme Equação 2.44, utilizando os quadrados dos resíduos, razão pela qual os maiores resíduos ficam com um peso maior em seu resultado.

$$RMSE = \left[\frac{\sum_{i=1}^H (Y_t - \hat{Y}_t)^2}{H} \right]^{\frac{1}{2}} \quad 2.44$$

2.2.6. Estudos relacionados

A primeira etapa desta Meta síntese consistiu na escolha da base de dados online a ser utilizada: Google Acadêmico. Após foi realizada a escolha das palavras-chave em inglês, para uma melhor abrangência de resultados: demanda de eletricidade (*electricity demand*), séries temporais (*time series*), modelos de previsão (*forecast models*). Esta procura foi realizada no título, resumo e palavras-chave de documentos do tipo “artigo”. Nos documentos selecionados foram escolhidos os documentos mais citados ou os documentos mais recentes que tivessem relação com o estudo desenvolvido, conforme Quadro 2.3

Quadro 2.3 – Quadro resumo dos estudos similares

Referência	Citação	Objetivo	Dados	Técnicas	Conclusões
Darberl ay e Slama; (2000)	428	Comparar se as previsões de modelos não lineares (Redes Neurais Artificiais) com os modelos lineares (do tipo ARMA).	De 1994 a 1995	ARMA RNA	A análise sugere que em algumas condições as redes neurais podem ser superiores aos modelos não lineares.
Erdogdu (2007)	444	Estimar um modelo de demanda de eletricidade da Turquia.	De 2000 a 2014	ARIMA	As projeções oficiais da demanda de eletricidade superestimam esta demanda, o que pode colocar em risco o desenvolvimento de uma política energética coerente.
Taylor <i>et al.</i> (2007)	667	Comparar métodos univariados para previsão usando dados da demanda intradiária de eletricidade de países europeus.	2005	PCA AR periódico SARMA Suavização Exponencial de Holt-Winters	O método que teve o melhor desempenho foi o de Suavização Exponencial de Holt-Winters com dupla sazonalidade.
Sumer <i>et al.</i> (2009)	63	Examinar as vantagens da previsão com os métodos ARIMA, SARIMA e com o modelo de regressão com variável latente sazonal.	De 1997 a 2005	ARIMA SARIMA Regressão com variável latente sazonal.	O modelo de regressão com variável latente sazonal encontra melhores resultados do que os modelos ARIMA e SARIMA.
Cruz <i>et al.</i> (2011)	146	Comparar empiricamente a precisão preditiva de um conjunto de métodos para previsão no mercado espanhol de eletricidade.	De 2007 a 2008	SARIMA Suavização Exponencial de Holt-Winters	O modelo de suavização exponencial para dupla sazonalidade apresentou de modo geral um desempenho melhor do que o ARIMA.
Henao <i>et al.</i> (2013)	14	Utilizar os modelos SARIMA e GSMN na modelagem da demanda energética.	De 1995 a 2009	SARIMA GSMN	O modelo híbrido proposto apresenta valores de RMSE e MAE inferiores aos modelos não híbridos em todos os três cenários estudados.
Yang <i>et al.</i> (2016)	151	Comparar os três métodos de previsão utilizando dados de energia do Estado de new South Wales, na Austrália.	2011	SARIMA BP ANFIS.	O modelo híbrido de maneira geral, levando em consideração as métricas RSME e MAE, apresentou melhores resultados que os modelos não híbridos.

Refe rência	Ci Ta ção	Objetivo	Dados	Técnicas	Conclusões
Bozkurt <i>et al.</i> (2017)	89	Prever a demanda energética do mercado Turco.	De 2013 a 2014	SARIMA RNA	O RNA se ajusta melhor que o SARIMA. Porém, o SARIMA tem um desempenho melhor nas previsões após feriados.
Sigauke (2017)	32	Modelar a demanda energética da África do Sul comparando os métodos de previsão.	De 2009 a 2013	GAM Lasso Benchmark	O modelo combinado se mostrou superior aos modelos individuais utilizados, apresentando valores de RMSE, MAE e MAPE inferiores.
Caro <i>et al.</i> (2020)	24	Estimar modelos para previsões de curto prazo de energia em cada uma das 24 horas do dia.	De 2014 a 2015	PARIMA	Os resultados indicam que o modelo pode ser utilizado em outros sistemas, com resultados bastante satisfatórios.
Andoh <i>et al.</i> (2021)	4	Explorar potenciais modelos de previsão da demanda de eletricidade para as regiões ocidentais de Gana.	De 2008 a 2020	SARIMA	Com um nível de confiança de 95%, os modelos desenvolvidos produziram resultados que podem ajudar a orientar na estimativa futura de eletricidade.

A previsão de demanda de eletricidade é uma área de pesquisa que tem recebido muita atenção nos últimos anos. É importante poder desenvolver, prever a sazonalidade e o crescimento da demanda e assim auxiliar as autoridades na tomada de decisão. Por se tratar de uma questão econômica, a precisão dos modelos é fator determinante. Por isso, vários autores discutem diferentes métodos de previsão, considerando vários modelos e técnicas ou mesmo modelos híbridos que buscam unificar os pontos fortes de cada técnica.

Darberllat e Slama (2000) realizam a previsão da demanda energética da República Tcheca utilizando os modelos ARMA e Redes Neurais Artificiais (RNA) com o objetivo de decidir se o problema da demanda elétrica da República Tcheca é realmente não linear e comparar se as previsões de modelos não lineares (Redes Neurais Artificiais) com os modelos lineares (do tipo ARMA). Constataram que a análise correlacional sugere que a previsão da evolução de curto prazo da carga elétrica Tcheca é principalmente um problema linear. Isso é confirmado pela comparação das previsões em algumas condições, onde as redes neurais podem ser superiores aos modelos não lineares.

Utilizando o modelo ARIMA, Erdogdu (2007) realiza a previsão da demanda energética da Turquia, com o objetivo de estimar um modelo de demanda de eletricidade e

obter estimativas de curto e longo prazo das elasticidades preço e renda. Como resultado obtido, verificou-se que a resposta dos consumidores às mudanças de preço e renda é bastante limitada e, portanto, há uma necessidade de regulação na economia do mercado turco. Conclui-se também que as atuais projeções oficiais de demanda de eletricidade superestimam muito a demanda de eletricidade, o que pode colocar em risco o desenvolvimento de uma política energética coerente em geral, e de um mercado de eletricidade saudável em particular.

Taylor *et al.* (2007), em seu estudo empírico, examinam a precisão de vários métodos univariados para a previsão da demanda de eletricidade de curto prazo. Entre os métodos examinados estão o método de suavização exponencial sugerido para dupla sazonalidade e um método baseado na Análise de Componentes Principais (PCA). Os métodos foram testados em uma variedade de países, incluindo Bélgica, Finlândia, França, Grã-Bretanha, Portugal, Irlanda, Itália, Noruega e Suécia. Os resultados foram insatisfatórios para o novo método de suavização exponencial e para o modelo AR periódico. Os métodos ARIMA e PCA funcionaram bem, mas o método que consistentemente teve o melhor desempenho foi o método de suavização exponencial dupla sazonal Holt-Winters.

Sumer *et al.* (2009) examinam as vantagens da previsão com os métodos ARIMA, SARIMA e com o modelo com variável latente sazonal entre si. Os resultados apontam que os modelos ARIMA e SARIMA não são bem-sucedidos na previsão da demanda de eletricidade. O modelo de regressão com variável latente sazonal dá resultados mais bem-sucedidos do que os modelos ARIMA e SARIMA, porque este modelo leva em consideração flutuações sazonais e quebras estruturais.

Cruz *et al.* (2011) também comparam empiricamente a precisão preditiva de um conjunto de métodos para previsão de preços à vista para o dia seguinte no mercado espanhol de eletricidade. Os métodos vêm de análise de séries temporais e disciplinas de inteligência artificial e incluem métodos univariados, multivariados, lineares e não lineares. Dentro dos métodos univariados, o ARIMA de dupla sazonalidade e a suavização exponencial para dupla sazonalidade são comparados e usados como benchmarks, e constatou-se que o modelo de suavização exponencial para dupla sazonalidade apresentou, de modo geral, um desempenho melhor do que o ARIMA. O último dos métodos testados são redes neurais feed-forward usadas como métodos de regressão não linear multivariada com recursos de aproximação de função universal. A influência da previsão de geração eólica na previsão de preços também é comprovada com esta abordagem

No âmbito da busca por melhores resultados, modelos combinados (híbridos) surgem como uma alternativa na modelagem da demanda energética, podendo apresentar acurácia

superior aos modelos individuais. Henao *et al.* (2013) propuseram uma abordagem de utilizar simultaneamente os modelos SARIMA e GSMN (Generalized Single Multiplicative Neuron) na modelagem da demanda energética. Sua proposta é que o modelo SARIMA trataria a parte linear dos dados, enquanto o GSMN trataria a parte não linear. Seu diferencial, portanto, é utilizar um modelo híbrido que otimize os parâmetros simultaneamente, utilizando o SARIMA e o GSMN, para superar os pontos negativos de não atingir os parâmetros ótimos e a configuração da rede neural por tentativa e erro.

Yang *et al.* (2016) propuseram um modelo que combina três métodos de modelagem e um método de otimização. Para a parte linear dos dados, utilizaram um modelo diff SARIMA (Difference Seasonal Autoregressive Integrated Moving Average), enquanto para a parte não linear dos dados, utilizaram os métodos BP (Back Propagation) e ANFIS (Adaptive Network-based Fuzzy Inference System). Com os resultados obtidos pelos três modelos, realizaram uma otimização dos parâmetros de pesos para cada resultado utilizando o algoritmo de otimização DE (Differential Evolution).

Bozkurt *et al.* (2017) compararam dois modelos para o mercado turco atual e apresentaram seus desempenhos. Foram usados os modelos de média móvel integrada autorregressiva sazonal (SARIMA) e a rede neural artificial (RNA). A principal contribuição deste estudo é a criação de modelos capazes de funcionar em tempo real e capazes de lidar com a natureza dinâmica do mercado desregulamentado. O autor concentra-se nessa questão fornecendo uma estimativa, previsão e comparação da demanda de eletricidade. A RNA se ajusta melhor que o SARIMA, mas o SARIMA faz melhores previsões após feriados.

Sigauke (2017) abordou a modelagem da demanda elétrica a médio prazo com resolução horária. Para isso, utilizou um modelo GAM (Generalized Additive Model) juntamente com interações de produtos de tensores. Utilizou ainda o método de regressão Lasso (Least Absolute Shrinkage and Selection Operator) para reduzir o número de variáveis trabalhadas. O modelo gerado foi então comparado com uma modelagem sem as interações de produtos de tensores e com um modelo que é usado como benchmark (Gradient Boosting Model). Após a comparação dos três modelos, denominados GAM-Lasso, GAM-te, Lasso e GBM, foi realizada a combinação dos modelos propostos por uma função pinball loss.

Caro *et al.* (2020) propõem uma abordagem para estimar modelos multiequacionais. Primeiramente, as equações de 24 horas são montadas para formar um modelo autorregressivo de média móvel periódica, o que melhora significativamente as previsões de curto prazo. Para reduzir o problema computacional, o modelo completo é estimado em duas etapas, e um modelo detalhado do efeito de temperatura não linear é incluído usando técnicas de regressão

spline. O método está sendo usado atualmente pelo Operador do Sistema de Transmissão Espanhol (Red Eléctrica de España, REE) para fazer previsões horárias de demanda de eletricidade de um a dez dias à frente.

Andoh *et al.* (2021) exploram potenciais modelos de séries temporais na previsão da demanda de eletricidade para as regiões ocidentais de Gana. Os dados secundários foram obtidos formalmente da sede regional da ECG para auxiliar no projeto de pesquisa a fim de estimar a quantidade de eletricidade necessária pelos consumidores da região. Isso foi feito usando o software ToolPak de análise de dados de séries temporais. Os resultados mostram que os modelos formulados são viáveis para previsões de consumo futuro e para orientar investimentos em projetos de fontes alternativas de energia para atender essas demandas futuras. Uma vez que existem padrões crescentes de demanda de energia na região, a flexibilidade dos modelos formulados pode ser muito útil para moldar políticas energéticas eficazes e eficientes.

Esses estudos demonstram a diversidade de abordagens e técnicas utilizadas na previsão de demanda de eletricidade e a importância contínua da pesquisa nesse campo para apoiar a tomada de decisões eficazes no setor de energia. Nota-se que os modelos híbridos, que combinam abordagens diferentes, como ARIMA e RNA, podem explorar o melhor de ambos os mundos, capturando tanto a linearidade quanto a não linearidade nos dados. Dentre as técnicas individuais, observa-se que a metodologia de Box-Jenkins é bastante usual e tem bons resultados preditivos.

3. METODOLOGIA

Neste capítulo, apresentam-se os procedimentos metodológicos utilizados na análise exploratória dos dados, bem como na aplicação da modelagem ARIMA Periódico (PARIMA) e ARIMA sazonal (SARIMA). Para a realização deste trabalho, foram utilizadas séries temporais com dados mensais, no período de janeiro de 2004 a dezembro de 2021, em que cada mês do ano, de janeiro a dezembro, possui 18 observações, totalizando 216 dados correspondentes à classe consumidora residencial de energia elétrica do estado do Espírito Santo. Todos os dados foram retirados da Empresa de Pesquisa Energética (EPE, 2020).

Os procedimentos descritos a seguir foram realizados com o auxílio dos softwares Gnu Regression Econometrics and Times-Series Library (Gretl, 2020), Rstudio (R CORE TEAM, 2020) com o auxílio das bibliotecas: *forecast* (Rob J Hyndman e Yeasmin Khandakar, 2008), *lmtest* (Achim Zeileis e Torsten Hothorn, 2002), *tseries* (Adrian Trapletti e Kurt Hornik, 2020), *finTS* (Spencer Graves, 2005), *ggplot2* (Hadley Wickham *et al.* 2016), *seasonal* (Christoph e Dirk Eddelbuettel, 2018) e *car* (John Fox and Sanford Weisberg, 2019) e da planilha Excel. Nas análises inferenciais foram utilizados 0,05 como nível de significância.

3.1. Procedimentos metodológicos na análise exploratória

Inicialmente, foi realizada uma análise exploratória dos dados da demanda de energia elétrica, utilizando métodos gráficos (diagrama em linhas, histograma e boxplot) e analíticos (medidas de tendência central, de dispersão e de forma da distribuição). Como os dados eram mensais, a série da demanda mensal de energia elétrica foi transformada em 12 subséries, uma para cada mês do ano. Esta análise também foi realizada em cada uma destas subséries, associadas a cada um dos 12 meses.

Em sequência, com o objetivo de avaliar a presença de sazonalidade, foi aplicada uma Análise de Variância paramétrica com blocos, com a utilização do teste F (Fisher, 1920), acompanhada de uma análise dos resíduos que empregou os testes de Shapiro-Wilk (Shapiro e Wilk, 1965), Levene (Levene, 1960) e Durbin-Watson (HILL, 2003) para verificar as propriedades de normalidade, homocedasticidade e independência, respectivamente. Em caso de não verificação destas suposições, foi utilizada uma Análise de Variância não paramétrica com o teste de Friedman (Skillings & Mac, 1981). A presença da sazonalidade indica a possibilidade de utilização dos métodos PARIMA e SARIMA.

Nesta análise exploratória, também foi avaliada a variabilidade da série e das subséries e, quando necessário, aplicada a transformação logarítmica com o objetivo de diminuí-la, de

forma a captar melhor o seu padrão de comportamento.

3.2. Procedimentos metodológicos na modelagem PARIMA

Nesta subseção apresenta-se os procedimentos abordados para a aplicação da modelagem ARIMA periódico ou PARIMA, sendo a análise descrita a seguir realizada individualmente em cada uma das 12 subséries.

Inicialmente realizou-se a divisão dos dados em dois períodos: ajuste, usado para captar este padrão de comportamento, correspondendo aos 15 primeiros anos e englobando cerca de 95% dos dados, e validação, usado para mensurar a capacidade preditiva, correspondendo ao último ano restante, englobando cerca de 5% dos dados.

Posteriormente, considerando apenas o período de ajuste, foram aplicados os testes ADF, KPSS e PP para avaliar a estacionariedade da série, sendo avaliados os resultados dos testes para modelos com tendência e sem tendência. Além disso, os testes também foram realizados utilizando as defasagens máximas determinadas pelos critérios propostos em Schwert (1989) e Hyndman e Athanasopoulos (2018). Em caso de divergência dos resultados encontrados, escolhe-se aquele que ocorre com maior frequência. Em caso de dúvida, foram examinadas as duas possibilidades. Depois de definir a ordem de integração da série, realiza-se duas modelagens, com e sem constante.

A partir da análise de estacionariedade da série, identificou-se a ordem dos parâmetros autorregressivos e de médias móveis. Nesta fase, foram construídos e analisados os gráficos da FAC e FACP para todos os meses, considerando todos os modelos gerados a partir da análise dos correlogramas, o que foi feito de acordo com critérios apresentados em Morettin e Toloi (2018). A partir dos valores encontrados para p e q , define-se o conjunto de modelos candidatos considerando todas as combinações possíveis entre $p \pm 1$ e $q \pm 1$.

Na sequência, são estimados os coeficientes dos modelos candidatos, o que foi feito pelo método da Máxima Verossimilhança. A parcimônia dos modelos gerados é medida pelos critérios de informação de Akaike (AIC) (AKAIKE, 1973), Akaike corrigido (AICc) (SUGIURA, 1978) e Bayesiano (BIC) (SCHWARZ, 1978).

Após, foi realizado o diagnóstico dos resíduos utilizando os testes de Ljung e Box (1978), para avaliar a presença de autocorrelação, o teste Jarque-Bera (1986), para verificar se os resíduos seguem uma distribuição Normal, e o teste de heterocedasticidade condicional autorregressiva (ARCH) (ENGLE, 1982), para verificar a presença de heterocedasticidade dos resíduos destes modelos candidatos.

Para validação e posterior determinação da capacidade preditiva, foram determinados os indicadores de qualidade: MAE, MAPE e RMSE. Caso haja divergências entre estas três

métricas, são considerados também os resultados do BIC e do RMSE, assim como os intervalos de confiança das previsões. Estas etapas estão sintetizadas na Figura 3.1.

Finalizando, depois de selecionar o modelo com melhor capacidade preditiva para cada mês, foi determinada uma nova capacidade preditiva para o conjunto de 12 modelos.

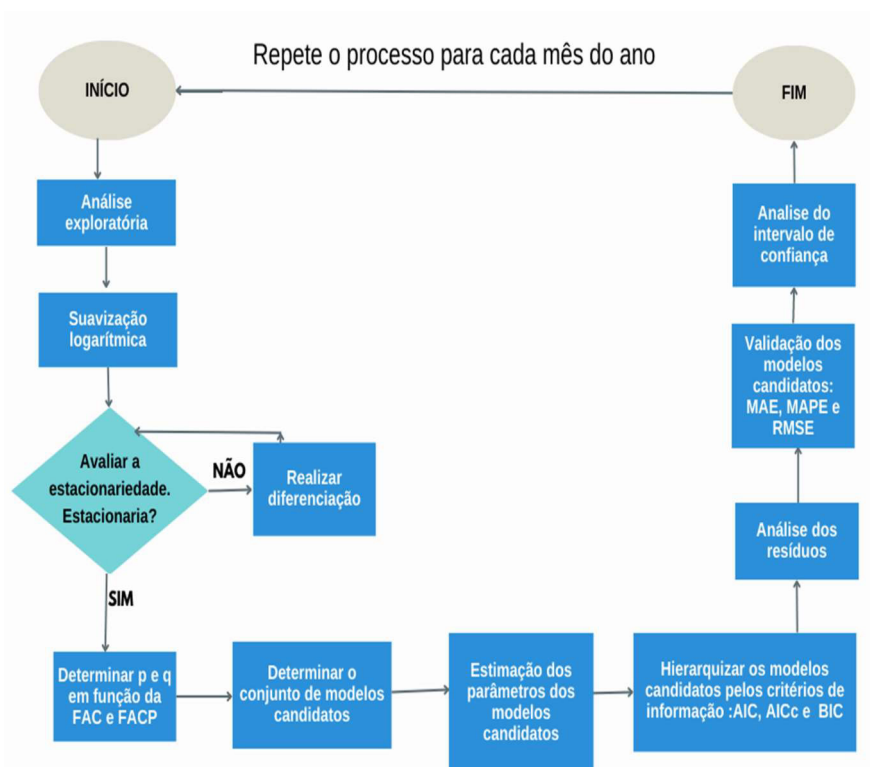


Figura 3.1- Fluxograma das etapas metodológicas para modelagem PARIMA

Fonte: Autora

3.3. Procedimentos metodológicos na modelagem SARIMA

Nesta seção apresenta-se os procedimentos metodológicos utilizados na aplicação da modelagem ARIMA sazonal ou SARIMA.

Inicialmente realizou-se a divisão da série em ajuste e validação, separando os primeiros 204 meses para fase de ajuste e 12 últimos para fase de validação. Posterior a divisão da série, foi verificada a estacionariedade das séries através dos testes ADF, KPSS e PP para a parte não sazonal e CH, HEGY e OCSB para a parte sazonal.

Na etapa 3 identificou-se a ordem dos parâmetros autorregressivos e de médias móveis da parte sazonal e não sazonal. Nesta fase foram construídos e analisados os gráficos da FAC e FACP. Como essas análises podem não apresentar com clareza os valores destes parâmetros, optou-se por atribuir os valores 0, 1, 2 e 3 a cada um dos parâmetros p, q, P e Q, considerando todas as 47 combinações possíveis entre estes. Resultaram 256 modelos e

através dos valores apresentados para o AIC, AICc e BIC foram considerados 5% do total, assim 13 modelos foram escolhidos de acordo com cada critério. sendo selecionados os modelos mais parcimoniosos

Já na etapa 5 foi realizado o diagnóstico dos resíduos obtidos com os modelos selecionados na etapa anterior. Utilizando os testes de Ljung-Box, Jarque-Bera e ARCH, foram verificadas a presença de autocorrelação, normalidade e heterocedasticidade dos resíduos destes modelos candidatos. Nesta etapa todos os modelos com resíduos autocorrelacionados foram descartados. Por fim, na etapa 6, realizou-se a validação dos modelos. Considerando os 12 últimos valores da amostra foram analisados os indicadores de qualidade: MAE, MAPE, RMSE. Com a escolha do modelo que melhor realiza previsões, foram construídos intervalos de confiança.

A Fig. 3.2 apresenta um esquema das etapas desenvolvidas para obtenção do modelo SARIMA.

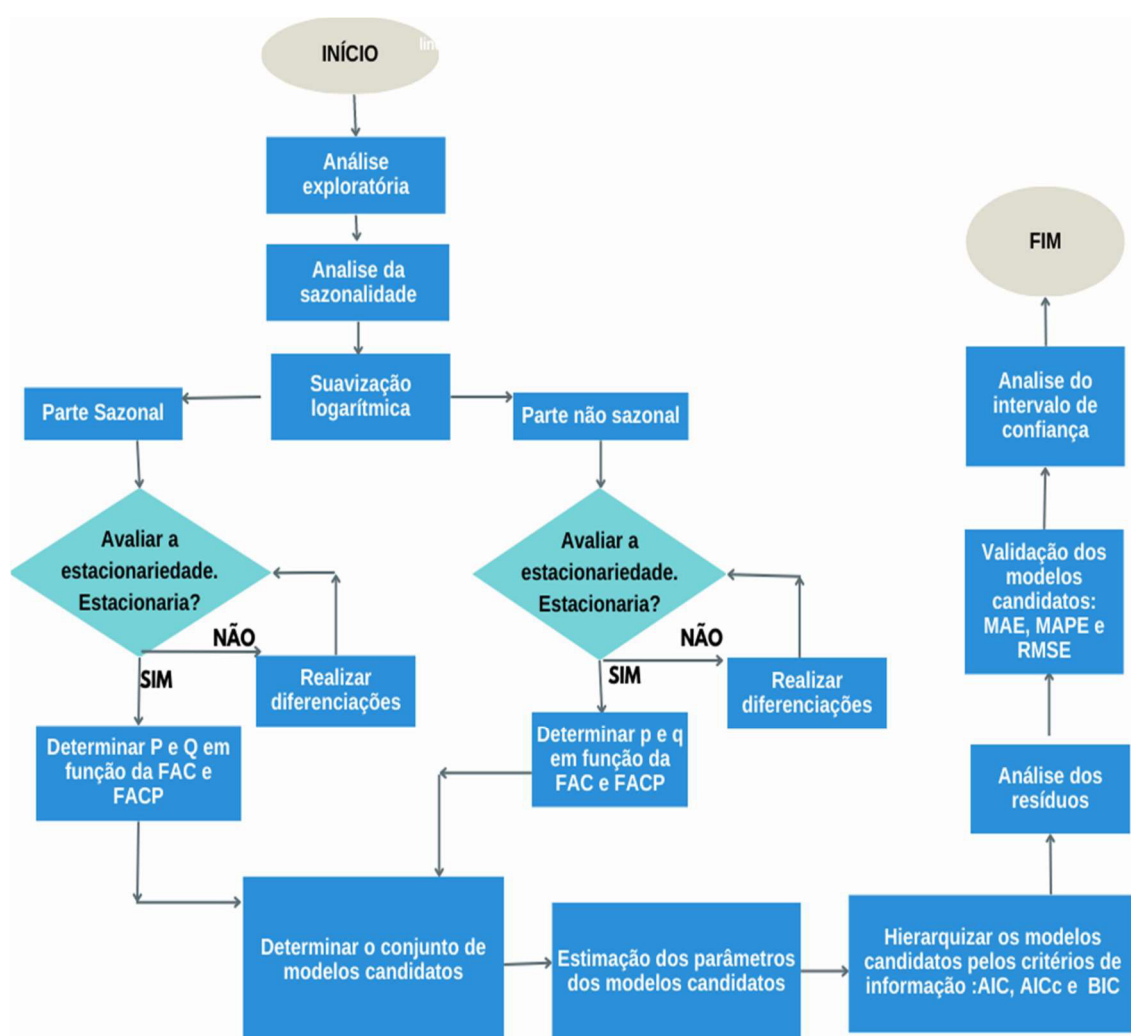


Figura 3.2- Fluxograma das etapas metodológicas para modelagem SARIMA

Fonte: Autora

4. RESULTADOS E DISCUSSÃO

Neste capítulo, são apresentados resultados e discussões da análise de dados da demanda de energia elétrica residencial no Espírito Santo. Na primeira seção, são mostrados os resultados da análise exploratória dos dados, na segunda seção, da aplicação do método PARIMA, enquanto na terceira seção, é apresentada a aplicação do SARIMA. Já na última seção, é feita uma discussão sobre o desempenho de algumas técnicas utilizadas nos modelos desenvolvidos.

4.1. Análise exploratória

A Figura 4.1 mostra o gráfico da série temporal da demanda mensal de energia elétrica para a classe residencial no estado do Espírito Santo, no período de 2004 até 2021. Observa-se que a série possui um crescimento ao longo dos anos, além de oscilações com picos em determinados períodos, o que sugere a presença de sazonalidade. Observa-se também que a variabilidade entre as oscilações aumenta de acordo com o aumento da demanda.

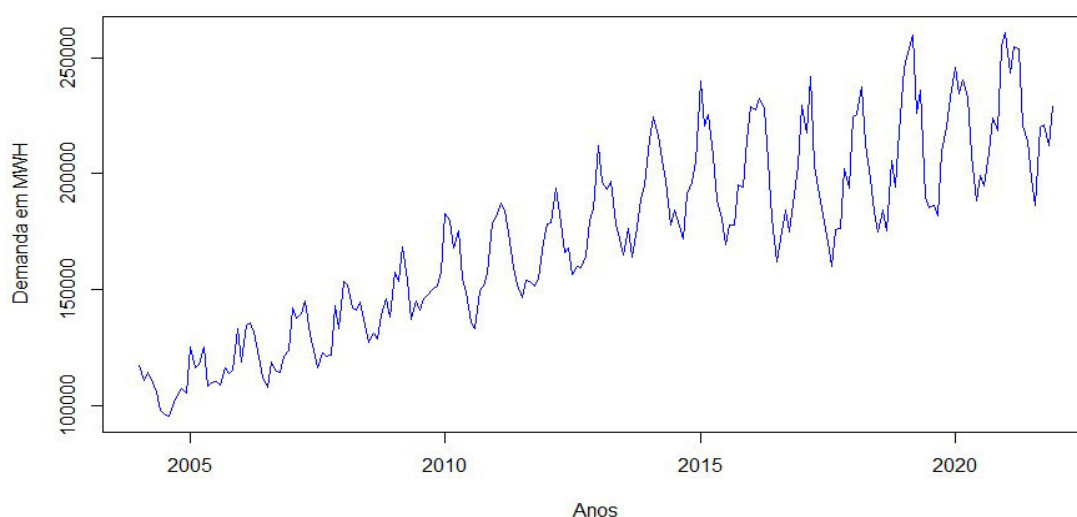


Figura 4.1-Gráfico em linhas para a demanda mensal de energia elétrica para a classe residencial/ ES, 2004-21

Fonte: Autora

Nos anos de 2020 e 2021 houve a pandemia de Covid-19 que afetou significativamente vários setores da economia, embora a Figura 4.1 sugira que a demanda de energia foi pouco afetada.

Para uma melhor análise visual das propriedades dos dados, foram construídos o Histograma e o Boxplot, apresentados na Figura 4.2. Observa-se que os dados não possuem outliers e são considerados simétricos.

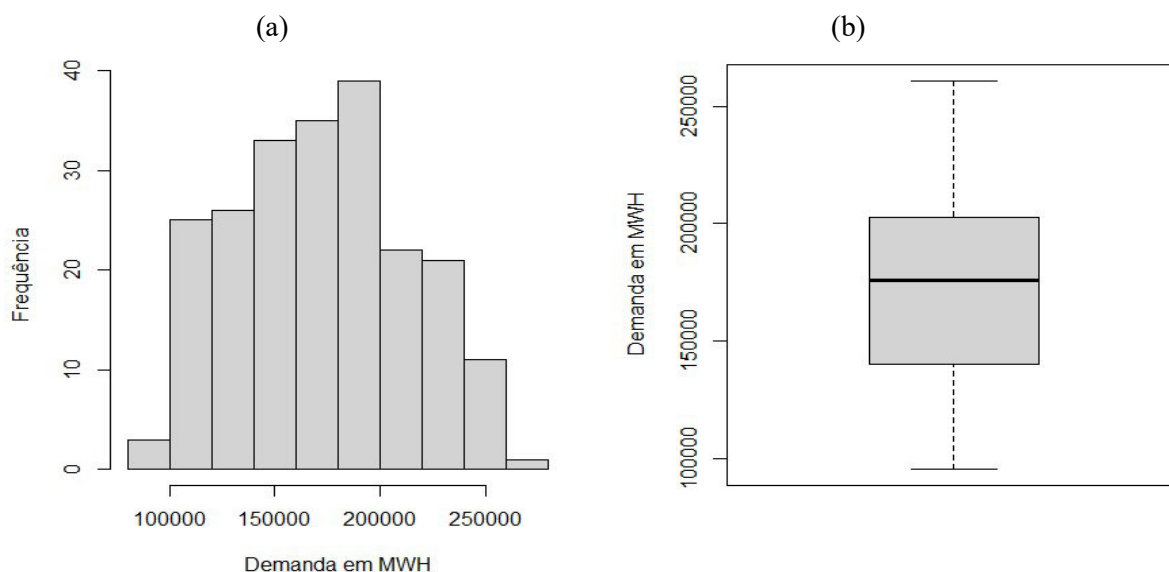


Figura 4.2- Histograma e gráfico de caixa para a demanda mensal de energia elétrica para a classe residencial/ ES, 2004-21

Fonte: Autora

Para avaliar melhor essas características foram construídas os *boxplots* para cada ano. Confirma-se, de acordo com a Figura 4.3, que ao longo dos anos existe um crescimento da demanda de energia elétrica, um aumento da variabilidade e ausência de *outliers*.

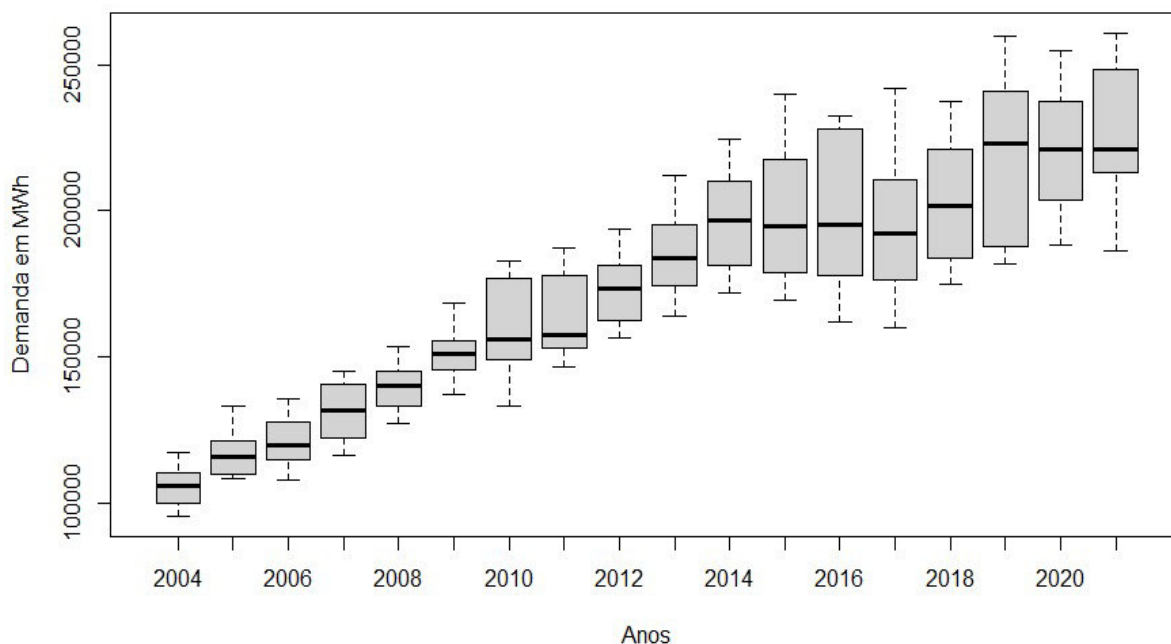


Figura 4.3- Gráfico de caixa para cada ano da demanda mensal de energia elétrica para a classe residencial/ ES, 2004- 21

Ainda para caracterizar o comportamento desta variável, foram calculadas algumas medidas descritivas que estão apresentadas na Tabela 4.1.

Tabela 4.1- Medidas resumo para a demanda mensal de energia elétrica para a classe residencial/ ES, 2004-21

Medidas	Resultados
Valor máximo (MWh)	260.910,00
Valor mínimo (MWh)	95.209,00
Média (MWh)	172.432,00
Mediana (MWh)	175.649,00
Desvio padrão (MWh)	40.740,76
Variância (MWh) ²	1.659.809.821,00
Coef. de assimetria	0,08
Coef. de curtose	-0,88
Coef. de variação (%)	23,63

Fonte: Autora

Pelas informações da Tabela 4.1, verifica-se, a partir dos valores mínimo e máximo, da variância e do desvio padrão, que houve uma variabilidade, de acordo com o coeficiente de variação os dados variaram 26,63% em torno do média. Através do coeficiente de curtose, verifica-se que a série temporal é platicúrtica, possuindo caudas leves. De acordo com o coeficiente de assimetria, confirma-se que a série é simétrica.

O gráfico da Figura 4.1 sugeriu que a série estudada tem um comportamento sazonal, razão pela qual esta propriedade foi avaliada de uma maneira mais formal. Uma Análise de Variância com blocos, e utilizando o teste F é empregada para testar as hipóteses nulas (H_0) de que não existe variação da média entre as observações dos tratamentos (meses) e nem entre as dos blocos (anos). Os resultados estão apresentados na Tabela 4.2.

Tabela 4.2- Resultado da ANOVA em blocos para a demanda mensal de energia elétrica para a classe residencial/ ES, 2004-21

Medidas	Fontes de variação		
	Blocos	Tratamentos	Erros
SQ	$2,96 \times 10^{11}$	$4,32 \times 10^{10}$	$1,77 \times 10^{10}$
QM	$1,74 \times 10^{10}$	$3,93 \times 10^9$	$9,48 \times 10^7$
Gl	17	11	187
F	183,67	41,48	
p-valor	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	

Fonte: Autora

Entretanto, de acordo com os resultados dos testes de Shapiro- Wilk, Levene e Durbin-Watson, apresentados na Tabela 4.3, a H_0 não é rejeitada ao nível de significância de 0,05 para o teste que avalia normalidade, não sendo encontradas evidências de que os resíduos não

apresentam distribuição normal. Por outro lado, rejeita-se H_0 para os testes que avaliam independência e homoscedasticidades, existindo evidências dos resíduos estarem autocorrelacionados e serem heterocedásticos.

Tabela 4.3- Análise dos resíduos ANOVA da demanda mensal de energia elétrica para a classe residencial/ ES, 2004- 21

Shapiro- Wilk		Levene		Durbin- Watson	
W	p-valor	F	p- valor	DW	p- valor
0,99	0,83	4,53	$<5,8 \times 10^{-8}$	0,86	$<2,2 \times 10^{-16}$

Fonte: Autora

Como na análise de resíduos não foram satisfeitas as suposições do modelo teórico utilizado, a Análise de Variância não paramétrica de Friedman foi aplicada. Os resultados estão apresentados na Tabela 4.4, onde é possível constatar que se rejeita H_0 a um nível de significância de 0,05. Assim, há evidências de que existe diferença significativa entre a demanda de pelo menos um dos meses em relação aos demais, confirmando a presença de sazonalidade.

Tabela 4.4- Resultado do teste de Friedman para a demanda mensal de energia elétrica para a classe residencial/ ES, 2004-21

Medidas	Valores
Q	157,30
Gl	11
p-valor	$2,2 \times 10^{-16}$

Fonte: Autora

As figuras 4.4 e 4.5 mostram gráficos em linhas e *boxplots* para cada mês, onde se verifica que, embora em todos os meses haja um crescimento da demanda de energia elétrica, no período 2004-21, o comportamento deste crescimento e sua variação foram diferentes em cada mês ao longo do período.

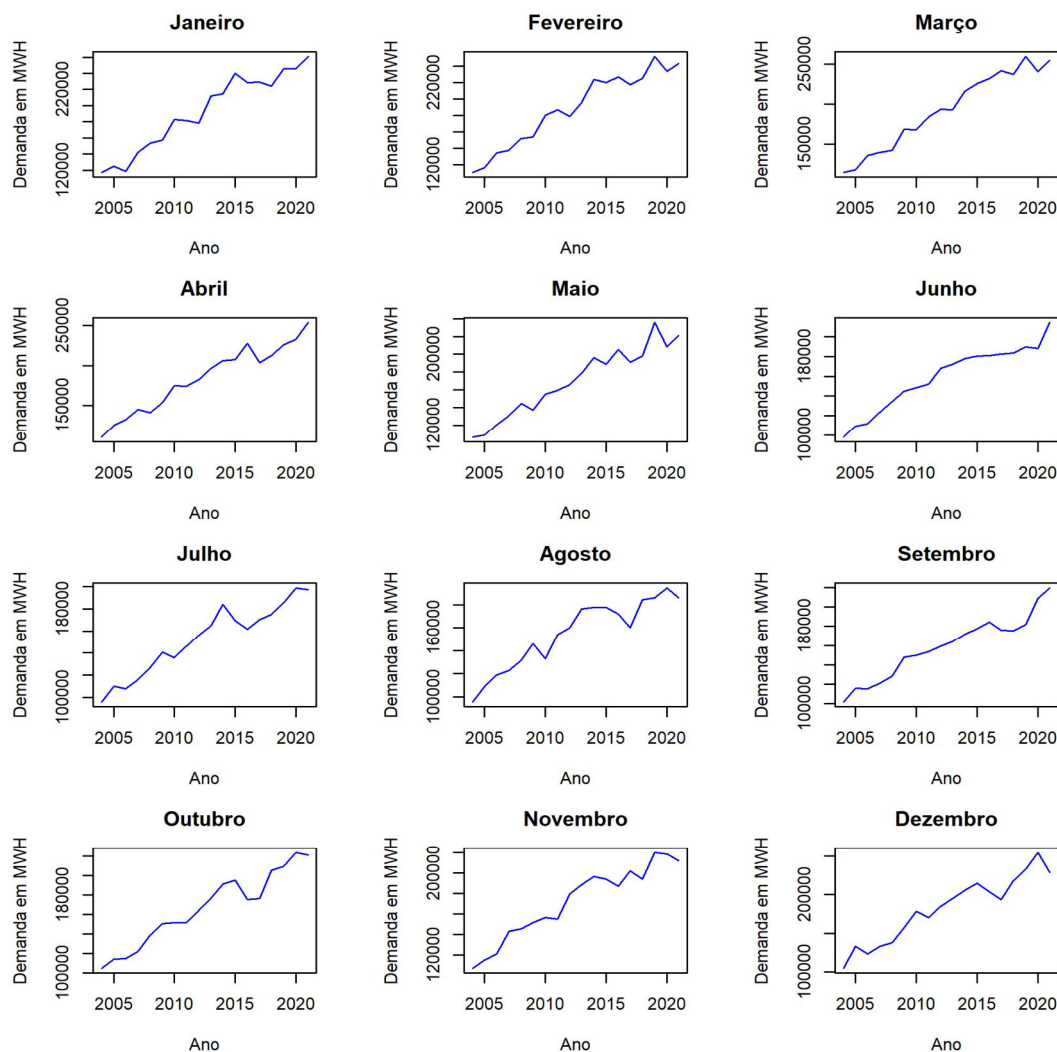


Figura 4.4- Gráfico em linhas para cada subsérie da demanda de energia elétrica residencial para os meses, ES/2004-21

Observa-se ainda, a partir da Figura 4.5, que a demanda de energia é menor para os meses julho, agosto e setembro, que apresentam temperaturas baixas, e maior para os meses de dezembro, janeiro e março, que apresentam temperaturas elevadas. Os gráficos desta figura confirmam ainda que não existem *outliers* nos dados tratados por mês.

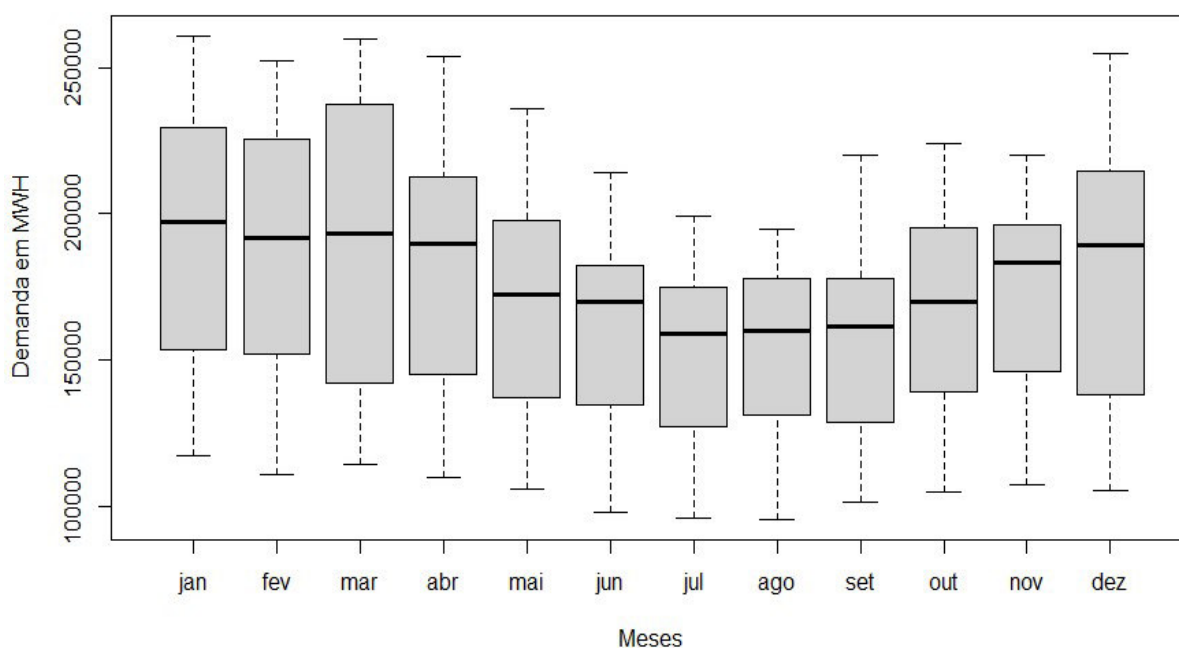


Figura 4.5- Gráfico de caixa para cada subsérie da demanda mensal de energia elétrica para a classe residencial/ ES, 2004- 21

Em sequência, a Tabela 4.5 mostra as medidas resumo para cada mês da demanda de energia elétrica no período 2004-21, onde é possível verificar que a maior demanda ocorre no mês de janeiro (altas temperaturas), e a menor demanda no mês de agosto (baixas temperaturas). Os coeficientes de curtose apontam que as séries são platicúrticas, com caudas leves, enquanto os coeficientes de assimetria apontam para uma assimetria leve, pois nenhum coeficiente atingiu o valor 0,5. Apesar dos coeficientes de curtose e assimetria serem bastantes próximos, existe uma pequena variação nessas medidas ao longo dos meses que pode ser observada através dos coeficientes de variação, percebe-se que os meses com maior variabilidade são março, janeiro e fevereiro.

Tabela 4.5- Medidas resumo para as doze subséries da demanda mensal de energia elétrica para a classe residencial/ ES, 2004-21

Meses	Valor Máximo (MWh)	Valor Mínimo (MWh)	Média (MWh)	Mediana (MWh)	Desvio padrão (MWh)	Coef. de assimetria	Coef. de curtose	Coef. de Variação (%)
Janeiro	260.910	117.210	192.052	197.336	47.622,00	-0,23	-1,49	24,79
Fevereiro	252.109	110.618	188.363	191.485	45.218,11	-0,29	-1,41	24,01
Março	259.827	114.342	192.400	193.242	48.449,92	-0,19	-1,50	25,15
Abril	253.879	109.482	183.653	189.476	41.747,59	-0,17	-1,27	22,74
Maio	235.668	105.932	169.395	172.123	39.050,10	-0,11	-1,29	23,05
Junho	214.127	97.920	158.887	169.986	32.603,44	-0,37	-1,10	20,55
Julho	198.960	95.662	152.372	159.035	31.742,41	-0,23	-1,28	20,86
Agosto	194.487	95.209	154.574	160.098	30.028,39	-0,45	-1,21	19,47
Setembro	219.804	101.358	158.456	161.576	32.637,92	-0,02	-0,95	20,60
Outubro	223.888	104.702	165.849	169.597	37.622,42	-0,07	-1,32	22,69
Novembro	220.135	107.157	171.501	183.287	35.660,91	-0,33	-1,27	20,82
Dezembro	254.796	105.104	181.677	189.225	42.375,66	-0,17	-1,19	23,31

Fonte: Autora

A análise exploratória realizada confirma as diferenças entre as medidas descritivas dos meses, assim como um padrão de crescimento diferente ao longo dos anos. Essas diferenças de características nas subséries correspondentes aos meses confirmam que a série possui sazonalidade, possibilitando a aplicação do método PARIMA e a aplicação do método SARIMA.

Nestas análises subsequentes, a série e as subséries analisadas foram divididas em período de ajuste ou treinamento, de 2004 até 2020, e período de validação ou teste, sendo composto pelo ano de 2021.

4.2. Modelagem PARIMA

Na modelagem PARIMA, a série analisada foi transformada em 12 subséries, uma para cada mês do ano, sendo a modelagem realizada separadamente. A Tabela 4.6 apresenta os resultados dos testes ADF, KPSS e PP, aplicados aos dados do período de ajuste, suavizados por meio da transformação logarítmica, para cada um dos doze meses.

A estacionariedade foi verificada para modelo com tendência e para modelo sem

tendência, considerando-se os valores dois (critério de Hyndman e Athanasopoulus) e sete (critério de Schwert) com defasagem máxima na aplicação do teste, sendo encontrados alguns resultados diferentes para diferentes testes e diferentes defasagens máximas, obtendo-se os resultados apresentados na Tabela 4.6.

Observando-se os resultados nesta tabela, nota-se uma unanimidade nos resultados dos três testes de raiz unitária para os meses de janeiro, fevereiro, abril e maio em que todos indicam que as séries se tornam estacionárias em primeira diferença, apresentando ordem de integração igual a um. Em contrapartida, o teste ADF divergiu dos resultados do KPSS e PP nos meses de julho, agosto, setembro, novembro e dezembro, sendo que o valor mais destoante para o grau de integração ocorreu quando adotado o critério de Schwert que definiu para a defasagem máxima o valor 7. Neste caso, o teste ADF e o teste PP apontaram a necessidade de 5 diferenças sucessivas para a obtenção da estacionariedade nos meses de setembro e outubro. De modo geral, o teste KPSS indicou que todos os meses teriam grau de integração 1, exceto para o mês de junho que, com defasagem máxima 7, indicou ordem de integração 2 para a remoção da raiz unitária, enquanto os testes ADF e PP apontaram que a série é estacionária em nível. Com defasagem máxima 2, os testes indicaram que a série é estacionária com grau de integração 1.

Tabela 4.6- Resultados dos testes de estacionariedade para cada subsérie da demanda mensal de energia elétrica/ES, 2004 – 21

Meses	Ordem de integração												OI
	ADF				KPSS				PP				
	dm = 2		dm = 7		dm = 2		dm = 7		dm = 2		dm = 7		
	T	C	T	C	T	C	T	C	T	C	T	C	
Janeiro	1	1	1	1	1	1	1	1	1	1	1	1	1
Fevereiro	1	1	1	1	1	1	1	1	1	1	1	1	1
Março	1	0	1	0	1	1	1	1	1	1	1	1	1
Abril	1	1	1	1	1	1	1	1	1	1	1	1	1
Maió	1	1	1	1	1	1	1	1	1	1	1	1	1
Junho	1	0	1	0	1	2	1	2	1	0	1	0	1
Julho	2	2	2	2	1	1	1	1	1	1	1	1	1
Agosto	2	2	2	2	1	1	1	1	1	1	1	1	2
Setembro	2	2	5	3	1	1	1	1	2	1	2	1	2
Outubro	1	1	1	1	1	1	1	1	2	1	5	1	1
Novembro	2	2	2	2	1	1	1	1	1	1	1	1	1
Dezembro	2	2	3	2	1	1	1	1	1	1	1	1	1

Obs: dm = defasagem máxima, T = Tendência, C = Constante, OI = ordem de integração adotada

Fonte: Autora.

As figuras de 4.6 até 4.17, apresentam os gráficos das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) para cada mês. Nesta etapa investigou-se a ordem da componente de média móvel (MA) e a ordem da componente autorregressiva (AR), o que foi feito nestes correlogramas, construídos levando em consideração as respectivas ordens de integração de cada série de acordo com os resultados da Tabela 4.6.

Nestes gráficos, verifica-se que apenas o mês de agosto apresenta correlação que extrapola o limite do intervalo construído com a confiança de 95% para estimá-la (Figura 4.13), sugerindo um modelo ARIMA(0,2,1). Para o mês de setembro o modelo sugerido é ARIMA(0,2,0), enquanto que para os demais meses é ARIMA(0,1,0).

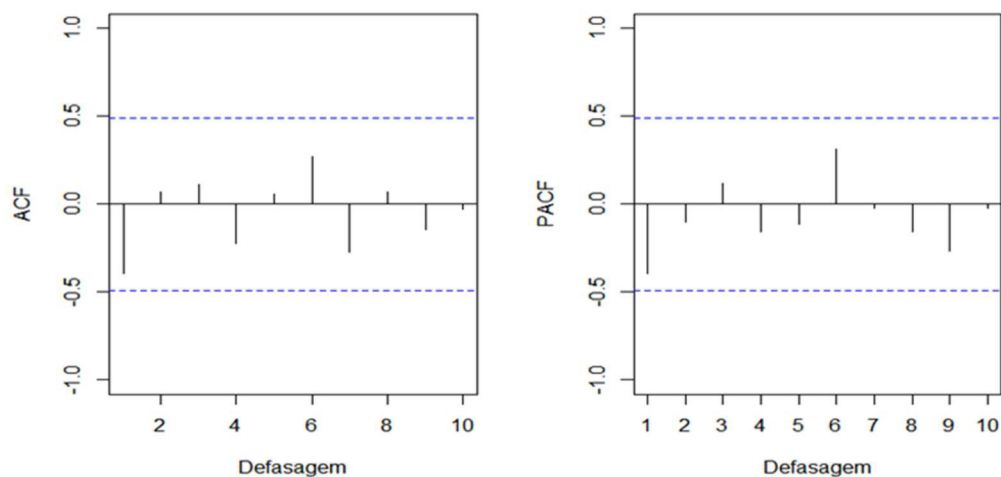


Figura 4.6- Gráfico da FAC e FACP para o mês de janeiro com grau de integração 1

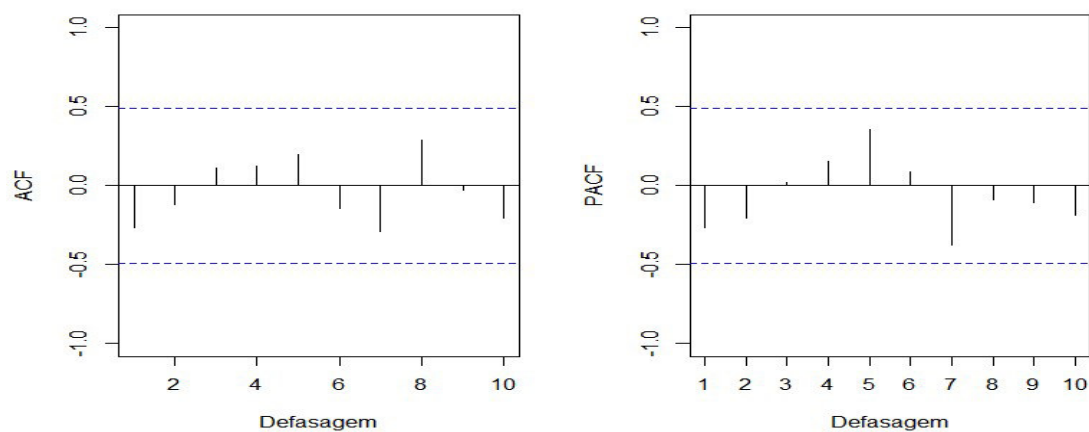


Figura 4.7- Gráfico da FAC e FACP para o mês de fevereiro com grau de integração 1

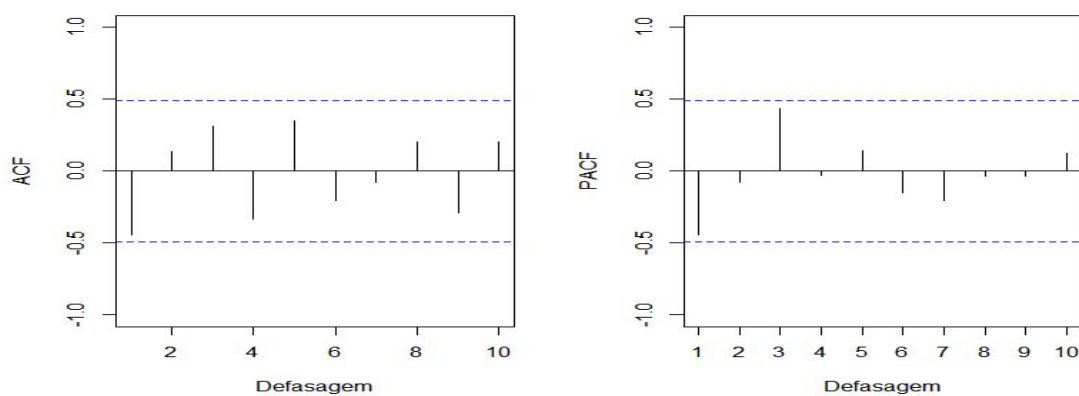


Figura 4.8- Gráfico da FAC e FACP para o mês de março com grau de integração 1

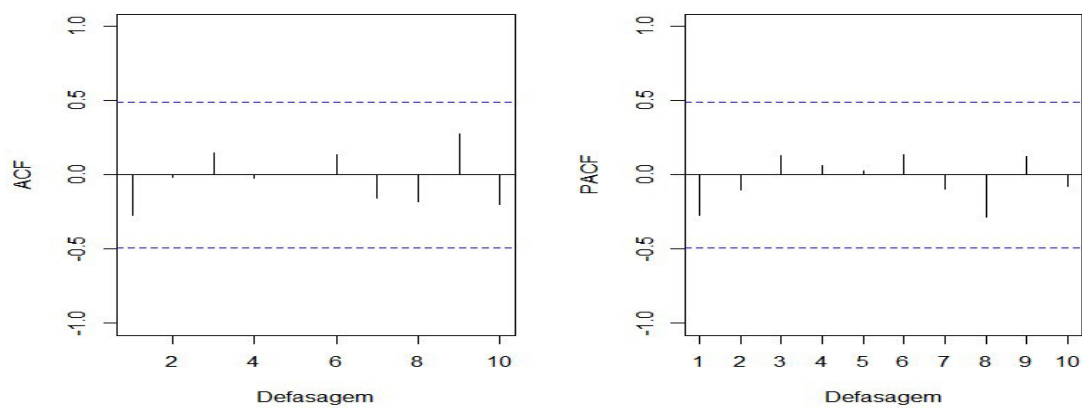


Figura 4.9- Gráfico da FAC e FACP para o mês de abril com grau de integração 1

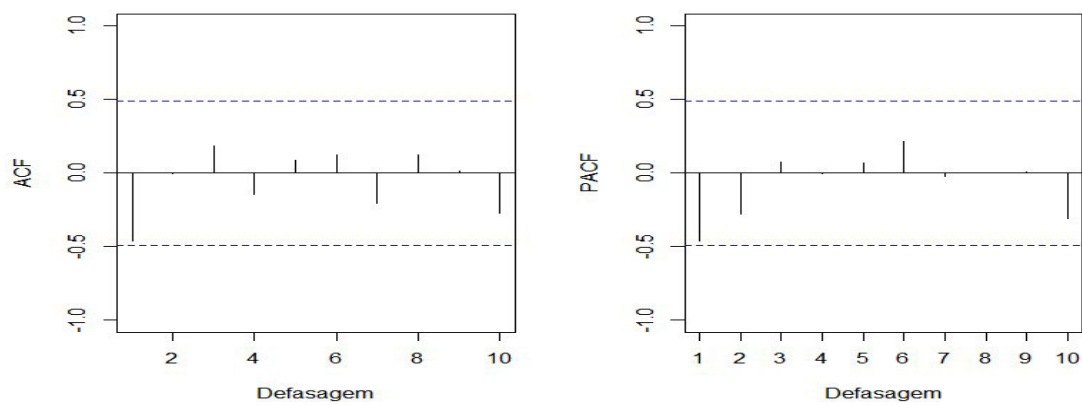


Figura 4.10- Gráfico da FAC e FACP para o mês de maio com grau de integração 1

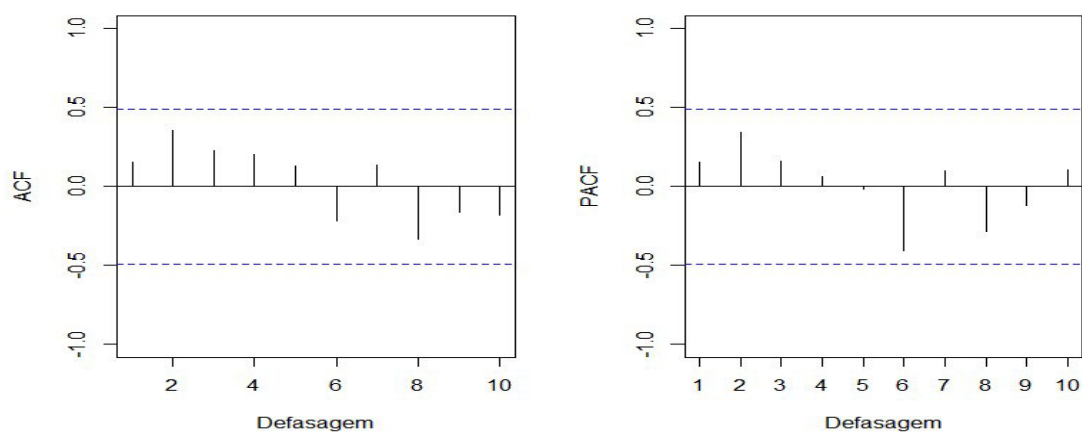


Figura 4.11- Gráfico da FAC e FACP para o mês de junho com grau de integração 1

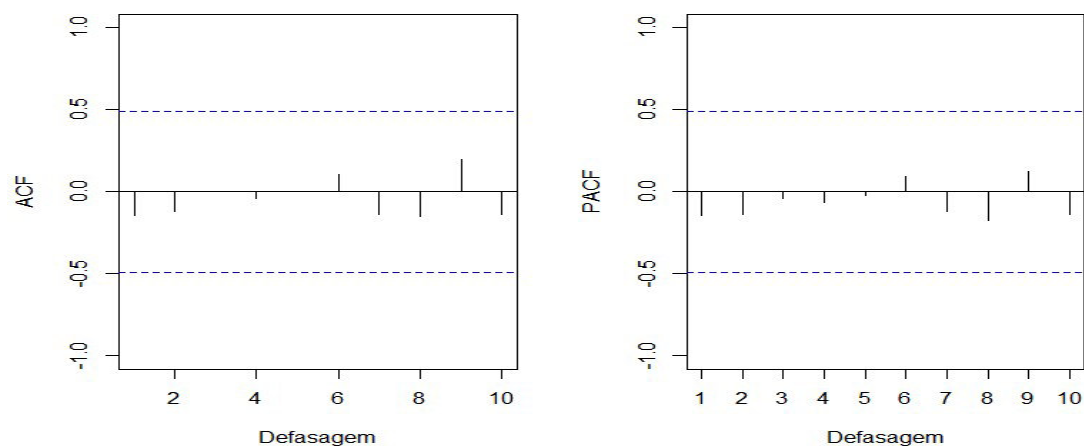


Figura 4.12- Gráfico da FAC e FACP para o mês de julho com grau de integração 1

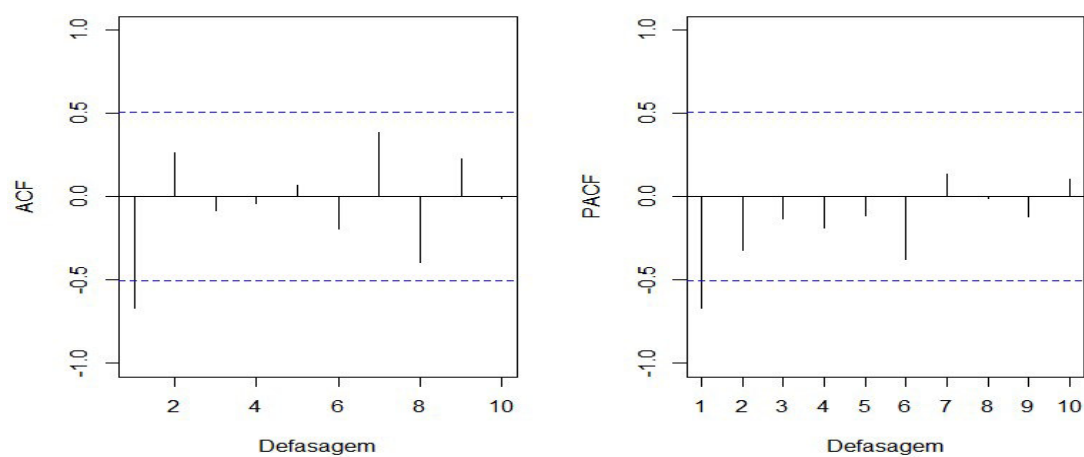


Figura 4.13- Gráfico da FAC e FACP para o mês de agosto com grau de integração 2

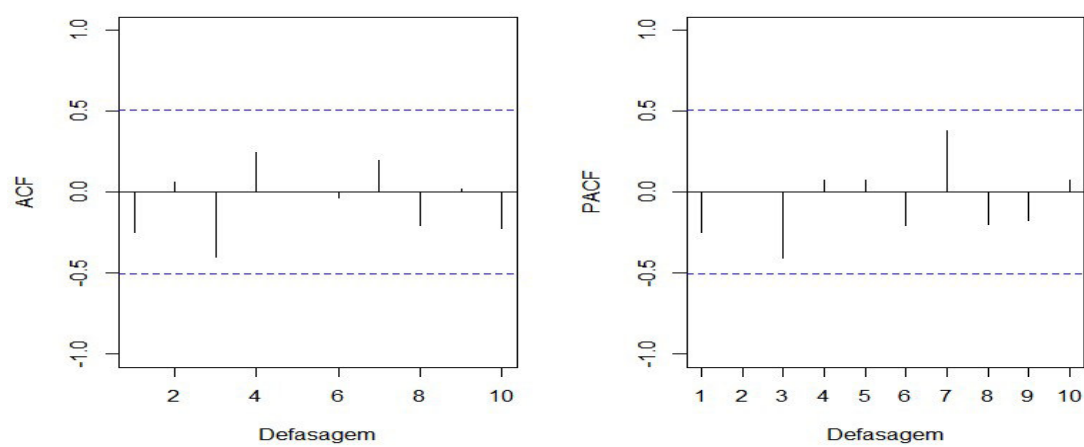


Figura 4.14- Gráfico da FAC e FACP para o mês de setembro com grau de integração 2

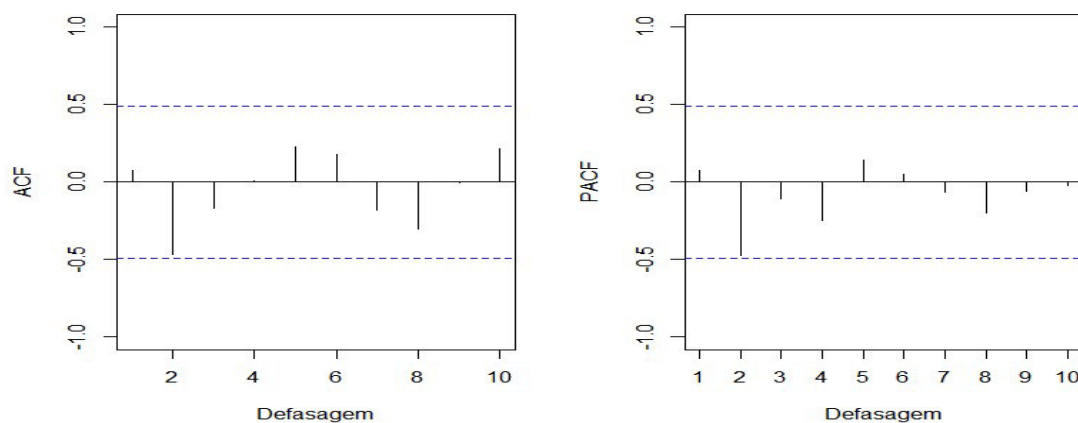


Figura 4.15- Gráfico da FAC e FACP para o mês de outubro com grau de integração 1

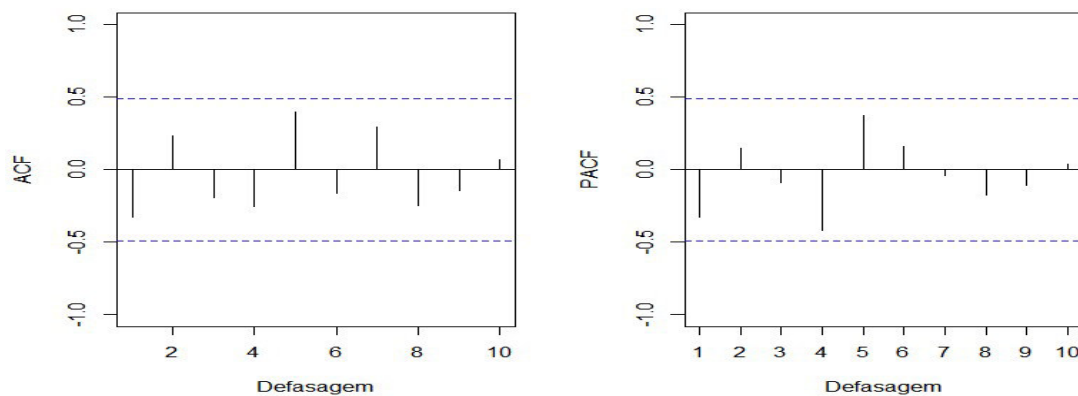


Figura 4.16- Gráfico da FAC e FACP para o mês de novembro com grau de integração 1

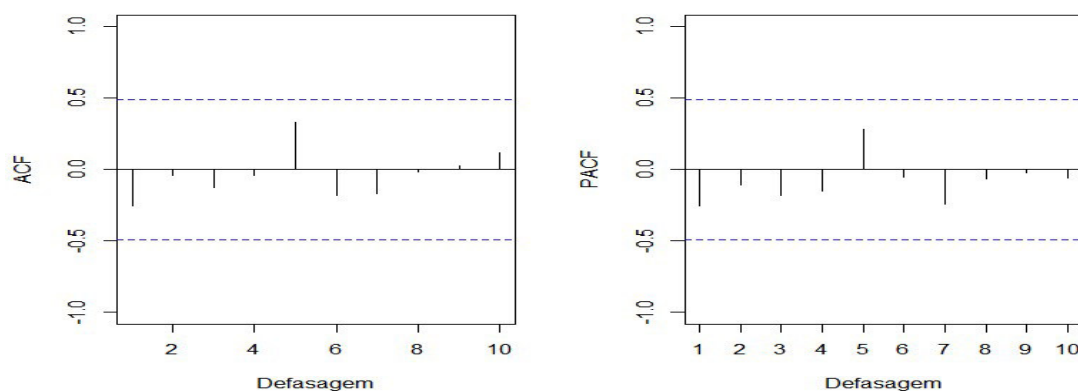


Figura 4.17- Gráfico da FAC e FACP para o mês de dezembro com grau de integração 1

Como a interpretação dos correlogramas apresenta um certo grau de subjetividade, definiram-se para cada mês um conjunto de modelos candidatos, o que foi feito somando e subtraindo 1 em p e q dos modelos sugeridos pelas interpretações dos gráficos da FAC e

FACP.

As tabelas de 4.7 a 4.18 mostram esses modelos candidatos para cada mês com os respectivos critérios de informação que medem a sua parcimonia, bem como o resultado da averiguação da presença de autocorrelação (Ljung-Box), normalidade (Jarque-Bera) e homocedasticidade (ARCH) nos resíduos.

Tabela 4.7- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de janeiro, ES, 2004-2020

Modelos gerados	AIC	AICc	BIC	Jarque-Bera (p-valor)	Ljung-Box (p-valor)	ARCH (p-valor)
ARIMA(0,1,0)*	-33,70	-32,77	-32,15	0,5361	0,3903	0,7254
ARIMA(1,1,0)	-34,34	-32,34	-32,03	0,5713	0,5978	0,7254
ARIMA(0,1,1)	-34,30	-32,30	-31,98	0,5361	0,6696	0,7254
ARIMA(1,1,1)	-32,44	-28,80	-29,35	0,5686	0,6366	0,7254

Obs: * identifica modelo considerado mais parcimonioso e que cumpre as suposições de independência, normalidade e homocedasticidade.

Tabela 4.8- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de fevereiro, ES, 2004-2020

Modelos gerados	AIC	AICc	BIC	Jarque-Bera (p-valor)	Ljung-Box (p-valor)	ARCH (p-valor)
ARIMA(0,1,0)*	-36,03	-35,11	-34,49	0,7246	0,3628	0,7254
ARIMA(0,1,1)	-35,82	-33,82	-33,51	0,7643	0,4762	0,7254
ARIMA(1,1,0)	-35,42	-33,42	-33,10	0,7372	0,3852	0,7254
ARIMA(1,1,1)	-33,83	-30,19	-30,74	0,7591	0,4727	0,7254

Obs: * identifica modelo considerado mais parcimonioso e que cumpre as suposições de independência, normalidade e homocedasticidade.

Tabela 4.9- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de março, ES, 2004-2020

Modelos gerados	AIC	AICc	BIC	Jarque-Bera (p-valor)	Ljung Box (p-valor)	ARCH (p-valor)
ARIMA(1,1,0)*	-42,71	-40,71	-40,39	0,9892	0,8046	0,7254
ARIMA(0,1,1)	-41,27	-39,27	-38,96	0,8984	0,3749	0,7254
ARIMA(0,1,0)	-39,95	-39,02	-38,40	0,9473	0,0147	0,7254
ARIMA(1,1,1)	-40,77	-37,14	-37,68	0,9807	0,8042	0,7254

Obs: * identifica modelo considerado mais parcimonioso e que cumpre as suposições de independência, normalidade e homoscedasticidade

Tabela 4.10- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de abril, ES, 2004-2020

Modelos gerados	AIC	AICc	BIC	Jarque-Bera (p-valor)	Ljung-Box (p-valor)	ARCH (p-valor)
ARIMA(0,1,0)*	-40,98	-40,06	-39,43	0,1171	0,4650	0,7254
ARIMA(0,1,1)	-40,44	-40,44	-38,13	0,2462	0,7892	0,7254
ARIMA(1,1,0)	-40,35	-38,35	-38,03	0,3369	0,7795	0,7254
ARIMA(1,1,1)	-38,48	-34,84	-35,39	0,2862	0,8057	0,7254

Obs: * identifica modelo considerado mais parcimonioso e que cumpre as suposições de independência, normalidade e homocedasticidade.

Tabela 4.11- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de maio, ES, 2004-2020

Modelos gerados	AIC	AICc	BIC	Jarque-Bera (p-valor)	Ljung-Box (p-valor)	ARCH (p-valor)
ARIMA(0,1,1)*	-35,72	-33,72	-33,40	0,5608	0,6169	0,7254
ARIMA(1,1,0)	-35,72	-33,72	-33,40	0,7224	0,5922	0,7254
ARIMA(1,1,1)	-34,52	-30,88	-31,43	0,6863	0,6517	0,7254
ARIMA(0,1,0)	-32,26	-31,33	-30,71	0,6029	0,3109	0,7254

Obs: * identifica modelo considerado mais parcimonioso e que cumpre as suposições de independência, normalidade e homocedasticidade.

Tabela 4.12- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de junho, ES, 2004-2020

Modelos gerados	AIC	AICc	BIC	Jarque-Bera (p-valor)	Ljung-Box (p-valor)	ARCH (p-valor)
ARIMA(0,1,0)*	-55,36	-54,44	-53,82	0,4075	0,1496	0,7254
ARIMA(1,1,0)	-53,88	-51,88	-51,56	0,3706	0,2455	0,7254
ARIMA(0,1,1)	-53,64	-51,64	-51,32	0,3819	0,2191	0,7254
ARIMA(1,1,1)	-51,92	-48,28	-48,83	0,4176	0,2944	0,7254

Obs: * identifica modelo considerado mais parcimonioso e que cumpre as suposições de independência, normalidade e homocedasticidade.

Tabela 4.13- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de julho, ES, 2004-2020

Modelos gerados	AIC	AICc	BIC	Jarque-Bera (p-valor)	Ljung-Box (p-valor)	ARCH (p-valor)
ARIMA(0,1,0)*	-40,04	-39,12	-38,49	0,4739	0,9219	0,7254
ARIMA(0,1,1)	-38,55	-36,55	-36,23	0,5897	0,9622	0,7254
ARIMA(1,1,0)	-38,41	-36,41	-36,09	0,5392	0,9491	0,7254
ARIMA(1,1,1)	-37,61	-33,97	-34,52	0,6781	0,9712	0,7254

Obs: * identifica modelo considerado mais parcimonioso e que cumpre as suposições de independência, normalidade e homoscedasticidade

Tabela 4.14- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de agosto, ES, 2004-2020

Modelos gerados	AIC	AICc	BIC	Jarque-Bera (p-valor)	Ljung-Box (p-valor)	ARCH (p-valor)
ARIMA(0,2,1)*	-29,54	-28,54	-28,13	0,4348	0,1233	0,7254
ARIMA(1,2,1)	28,62	-26,44	-26,50	0,8724	0,4912	0,7254
ARIMA(0,2,2)	-28,48	-26,30	-26,36	0,8462	0,4010	0,7254
ARIMA(1,2,0)	-27,30	-26,30	-25,88	0,4626	0,4149	0,7254
ARIMA(2,2,0)	-27,02	-24,84	-24,89	0,7885	0,6681	0,7254
ARIMA(2,2,1)	-26,65	-22,65	-23,82	0,8953	0,5065	0,7254
ARIMA(1,2,2)	-26,65	-22,65	-23,81	0,8896	0,5042	0,7254
ARIMA(2,2,2)	-24,65	-17,98	-21,11	0,8946	0,5062	0,7254
ARIMA(0,2,0)	-21,04	-20,73	-20,33	0,4348	0,0045	0,7254

Obs: * identifica modelo considerado mais parcimonioso e que cumpre as suposições de independência, normalidade e homocedasticidade.

Tabela 4. 15- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de setembro, ES, 2004-2020

Modelos gerados	AIC	AICc	BIC	Jarque-Bera (p-valor)	Ljung-Box (p-valor)	ARCH (p-valor)
ARIMA(0,2,1)*	-38,99	-37,99	-37,58	0,4148	0,4132	0,7254
ARIMA(0,2,0)	-36,52	-36,22	-35,82	0,4148	0,3696	0,7254
ARIMA(1,2,0)	-36,09	-35,09	-34,68	0,9495	0,5034	0,7254
ARIMA(1,2,1)	-34,80	-32,62	-32,68	0,7633	0,4683	0,7254

Obs: * identifica modelo considerado mais parcimonioso e que cumpre as suposições de independência, normalidade e homocedasticidade.

Tabela 4.16- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de outubro, ES, 2004-2020

Modelos gerados	AIC	AICc	BIC	Jarque-Bera (p-valor)	Ljung-Box (p-valor)	ARCH (p-valor)
ARIMA(0,1,1)*	-37,72	-36,79	-36,17	0,5870	0,7647	0,7254
ARIMA(1,1,0)	-35,68	-34,76	-34,13	0,8491	0,3324	0,7254
ARIMA(1,1,1)	-35,92	-33,92	-33,6	0,5952	0,8198	0,7254
ARIMA(0,1,0)	-34,98	-34,69	-34,21	0,4763	0,2193	0,7254

Obs: * identifica modelo considerado mais parcimonioso e que cumpre as suposições de independência, normalidade e homocedasticidade.

Tabela 4. 17- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de novembro, ES, 2004-2020

Modelos gerados	AIC	AICc	BIC	Jarque-Bera (p-valor)	Ljung-Box (p-valor)	ARCH (p-valor)
ARIMA(0,1,0)*	-40,55	-39,62	-39,00	0,6367	0,0971	0,7254
ARIMA(1,1,0)	40,32	38,32	-38,01	0,9227	0,3926	0,7254
ARIMA(0,1,1)	-39,92	37,92	37,60	0,9064	0,3806	0,7254
ARIMA(1,1,1)	-38,69	-35,05	-35,60	0,9166	0,4117	0,7254

Obs * identifica modelo considerado mais parcimonioso e que cumpre as suposições de independência, normalidade e homocedasticidade.

Tabela 4.18- Resultados dos critérios de informação e da análise de resíduos dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de dezembro, ES, 2004-2020

Modelos gerados	AIC	AICc	BIC	Jarque-Bera (p-valor)	Ljung-Box (p-valor)	ARCH (p-valor)
ARIMA(0,1,0)*	-27,38	-27,09	-26,61	0,9116	0,6792	0,7254
ARIMA(0,1,1)	-25,43	-24,51	-23,89	0,9116	0,6429	0,7254
ARIMA(1,1,0)	-25,46	-24,54	-23,92	0,9554	0,6161	0,7254
ARIMA(1,1,1)	-23,40	-21,40	-21,08	0,8980	0,6788	0,7254

Obs: * identifica modelo considerado mais parcimonioso e que cumpre as suposições de independência, normalidade e homocedasticidade.

Avaliando se os modelos satisfazem o pressuposto de que os resíduos são independentes, considerando um nível de significância de 0,01, constatou-se que o modelo ARIMA(0,2,0) para o mês de agosto não atendeu ao critério, sendo eliminado. Todos os

outros modelos obtiveram a hipótese nula não rejeitada, indicando que não foram encontradas evidências de autocorrelação entre os resíduos.

De acordo com o teste de ARCH, os resíduos de todos os modelos não apresentaram evidências de possuir heterocedasticidade, pois a hipótese nula para esse teste não pode ser rejeitada, o mesmo acontecendo com o teste de Jarque-Bera, não apresentando evidências de não normalidade nos resíduos dos modelos candidatos das subséries.

Para cada subsérie, entre os modelos candidatos, apenas aqueles que não apresentam evidências da presença de autocorrelação em seus resíduos tiveram sua capacidade preditiva analisada. As tabelas de 4.19 até 4.30 apresentam as medidas de acurácia: raiz do erro quadrático médio (RMSE), o erro absoluto médio (MAE), o erro absoluto percentual médio (MAPE), dos modelos selecionados, sendo identificado aquele que apresenta melhor capacidade preditiva.

Tabela 4.19- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de janeiro, ES, 2021

Modelos ARIMA	MAE (MWh)	MAPE (%)	RMSE (MWh)
ARIMA(1,1,1)*	1174,47	0,4501	1174,47
ARIMA(0,1,1)	1290,31	0,4945	1290,31
ARIMA(1,1,0)	1301,11	0,4986	1301,11
ARIMA(0,1,0)	3297,77	1,2630	3297,77

Obs:* modelo com melhor capacidade preditiva.

Tabela 4.20- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de fevereiro, ES, 2021

Modelos ARIMA	MAE (MWh)	MAPE (%)	RMSE (MWh)
ARIMA(0,1,0)*	2433,26	1,0008	2433,26
ARIMA(1,1,0)	12341,04	5,0762	12341,04
ARIMA(0,1,1)	12371,66	5,0888	12371,66
ARIMA(1,1,1)	12652,45	5,2043	12652,45

Obs: * modelo com melhor capacidade preditiva.

Tabela 4.21- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de março, ES, 2021

Modelos ARIMA	MAE (MWh)	MAPE (%)	RMSE (MWh)
ARIMA(0,1,0)*	2773,18	1,0897	2773,18
ARIMA(0,1,1)	10945,59	4,3009	10945,59
ARIMA(1,1,0)	17186,23	6,7532	17186,23
ARIMA(1,1,1)	17668,30	6,9426	17668,30

Obs: * modelo com melhor capacidade preditiva.

Tabela 4.22- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de abril, ES, 2021

Modelos ARIMA	MAE (MWh)	MAPE (%)	RMSE (MWh)
ARIMA(1,1,0)*	8748,90	3,4460	8748,90
ARIMA(0,1,1)	8857,96	3,4890	8857,96
ARIMA(1,1,1)	8881,16	3,4981	8881,16
ARIMA(0,1,0)	9684,94	3,8147	9684,94

Obs: * modelo com melhor capacidade preditiva.

Tabela 4.23- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de maio, ES, 2021

Modelos ARIMA	MAE (MWh)	MAPE (%)	RMSE (MWh)
ARIMA(0,1,0)*	3731,90	1,6908	3731,90
ARIMA(0,1,1)	13144,48	5,9553	13144,48
ARIMA(1,1,1)	19401,75	8,7903	19401,75
ARIMA(1,1,0)	21471,18	9,7279	21471,18

Obs: * modelo com melhor capacidade preditiva.

Tabela 4.24- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de junho, ES, 2021

Modelos ARIMA	MAE (MWh)	MAPE (%)	RMSE (MWh)
ARIMA(0,1,0)*	18170,01	8,4856	18170,01
ARIMA(0,1,1))	19224,57	8,9781	19224,57
ARIMA(1,1,0)	20107,92	9,3906	20107,92
ARIMA(1,1,1)	20094,02	9,3841	20094,02

Obs: * modelo com melhor capacidade preditiva.

Tabela 4.25- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de julho, ES, 2021

Modelos ARIMA	MAE (MWh)	MAPE (%)	RMSE (MWh)
ARIMA(0,1,1)*	9191,77	4,6548	9191,77
ARIMA(1,1,0)	9702,56	4,9134	9702,56
ARIMA(0,1,0)	10809,30	5,4739	10809,30
ARIMA(1,1,1)	12129,97	6,1427	12129,97

Obs: * modelo com melhor capacidade preditiva.

Tabela 4.26- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de agosto, ES, 2021

Modelos ARIMA	MAE (MWh)	MAPE (%)	RMSE (MWh)
ARIMA(1,2,0)*	13099,86	7,0441	13099,86
ARIMA(0,2,2)	13467,13	7,2416	13467,13
ARIMA(1,2,2)	13857,49	7,4515	13857,49
ARIMA(1,2,1)	13980,10	7,5174	13980,10
ARIMA(2,2,2)	13988,03	7,5217	13988,03
ARIMA(2,2,1)	14037,44	7,5482	14037,44
ARIMA(0,2,1)	16480,32	8,8618	16480,32
ARIMA(2,2,0)	20281,02	10,9056	20281,02

Obs: * modelo com melhor capacidade preditiva.

Tabela 4.27- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de setembro, ES, 2021

Modelos ARIMA	MAE (MWh)	MAPE (%)	RMSE (MWh)
ARIMA(0,2,1)*	1193,78	0,5431	1193,78
ARIMA(1,2,0)	10732,72	4,8828	10732,72
ARIMA(1,2,1)	12884,71	5,8619	12884,71
ARIMA(0,2,0)	20584,09	9,3647	20584,09

Obs: * modelo com melhor capacidade preditiva.

Tabela 4.28- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de outubro, ES, 2021

Modelos ARIMA	MAE (MWh)	MAPE (%)	RMSE (MWh)
ARIMA(0,1,0)*	3036,10	1,3747	3036,10
ARIMA(1,1,0)	9188,53	4,1604	9188,53
ARIMA(1,1,1)	23269,19	10,5361	23269,19
ARIMA(0,1,1)	25110,10	11,3696	25110,10

Obs: * modelo com melhor capacidade preditiva.

Tabela 4.29- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de novembro, ES, 2021

Modelos ARIMA	MAE (MWh)	MAPE (%)	RMSE (MWh)
ARIMA(0,1,0)*	16411,28	7,7431	16411,28
ARIMA(0,1,1)	18666,36	8,8071	18666,36
ARIMA(1,1,0)	20431,35	9,6399	20431,35
ARIMA(1,1,1)	22717,95	10,7187	22717,95

Obs: * modelo com melhor capacidade preditiva.

Tabela 4.30- Capacidade preditiva dos modelos candidatos gerados para a demanda mensal de energia elétrica da classe residencial no mês de dezembro, ES, 2021

Modelos ARIMA	MAE (MWh)	MAPE (%)	RMSE (MWh)
ARIMA(1,1,1)*	25158,98	10,9998	25158,98
ARIMA(0,1,0)	26075,10	11,4003	26075,10
ARIMA(0,1,1)	27301,15	11,9364	27301,15
ARIMA(1,1,0)	28183,55	12,3222	28183,55

Obs: * modelo com melhor capacidade preditiva.

Os modelos identificados como apresentando melhor capacidade preditiva em cada subsérie estão apresentados na Tabela 4.31. O melhor desempenho na previsão foi obtido pelos modelos encontrados para os meses de janeiro e setembro, obtendo um MAPE inferior a 0,6%. Vale ressaltar que nos meses de setembro e agosto houve discordância nos testes que avaliam a estacionariedade, razão pela qual foram estimados modelos considerando grau de integração 1 e 2. Apesar dos modelos terem conseguido melhores previsões quando construídos com grau de integração 2, obtiveram piores resultados para os critérios AIC, AICc e BIC, quando comparados com os modelos com grau de integração 1.

Todas as subséries foram modeladas visando identificar o modelo com melhor capacidade preditiva, sendo consideradas diferentes estruturas desde que as suposições

teóricas fossem verificadas. Os meses de janeiro, fevereiro, março, abril, maio, junho, julho e novembro foram modelados considerando a inclusão de uma constante. Entretanto, os meses de agosto, setembro, outubro e dezembro foram modelados sem a inclusão de constante.

Em sequência, foi determinada a capacidade preditiva da modelagem PARIMA, formada pelos melhores modelos de previsão em cada subsérie. Os resultados estão apresentados na Tabela 4.31 e no gráfico da Figura 4.18, sendo a modelagem considerada satisfatória, pois obteve um RMSE de 11563,78 MWh, MAE de 8760,29 MWh e MAPE de 4,0435%.

Tabela 4.31- Capacidade preditiva do Modelo PARIMA da demanda mensal de energia elétrica para a classe residencial e respectivas medidas de acurácia, ES/2021

Meses	Modelos selecionados	MAE (MWh)	MAPE (%)	RMSE (MWh)
Janeiro	ARIMA(1,1,1)	1174,47	0,4501	1174,47
Fevereiro	ARIMA(0,1,0)	2433,26	1,0008	2433,26
Março	ARIMA(0,1,0)	2773,18	1,0897	2773,18
Abril	ARIMA(1,1,0)	8748,90	3,4460	8748,90
Maio	ARIMA(0,1,0)	3731,90	1,6908	3731,90
Junho	ARIMA(0,1,0)	18170,01	8,4856	18170,01
Julho	ARIMA(0,1,1)	9191,77	4,6548	9191,77
Agosto	ARIMA(1,2,0)	13099,86	7,0441	13099,86
Setembro	ARIMA(0,2,1)	1193,78	0,5431	1193,78
Outubro	ARIMA(0,1,0)	3036,10	1,3747	3036,10
Novembro	ARIMA(0,1,0)	16411,28	7,7431	16411,28
Dezembro	ARIMA(1,1,1)	25158,98	10,9998	25158,98
Modelagem PARIMA		8760,29	4,0435	11563,78

Fonte: Autora

De acordo com os resultados encontrados na Tabela 4.31, observa-se que alguns modelos das subséries têm melhor capacidade preditiva do que outros, já que os conjuntos de valores de RMSE e MAE variam em uma amplitude de 24.784,79 MWh e 10,5491% para o MAPE. Os meses de janeiro e dezembro correspondem aos meses que obtiveram, respectivamente, a melhor e a pior previsão. Aproximadamente 41% dos modelos classificados como os que produziram melhores previsões em cada mês também indicaram um melhor resultado para os critérios de informação, e em 50% dos casos o MAPE encontrou um resultado inferior a 2,5%. A Figura 4.18 mostra o gráfico de previsão do modelo

PARIMA de acordo com os modelos selecionados.

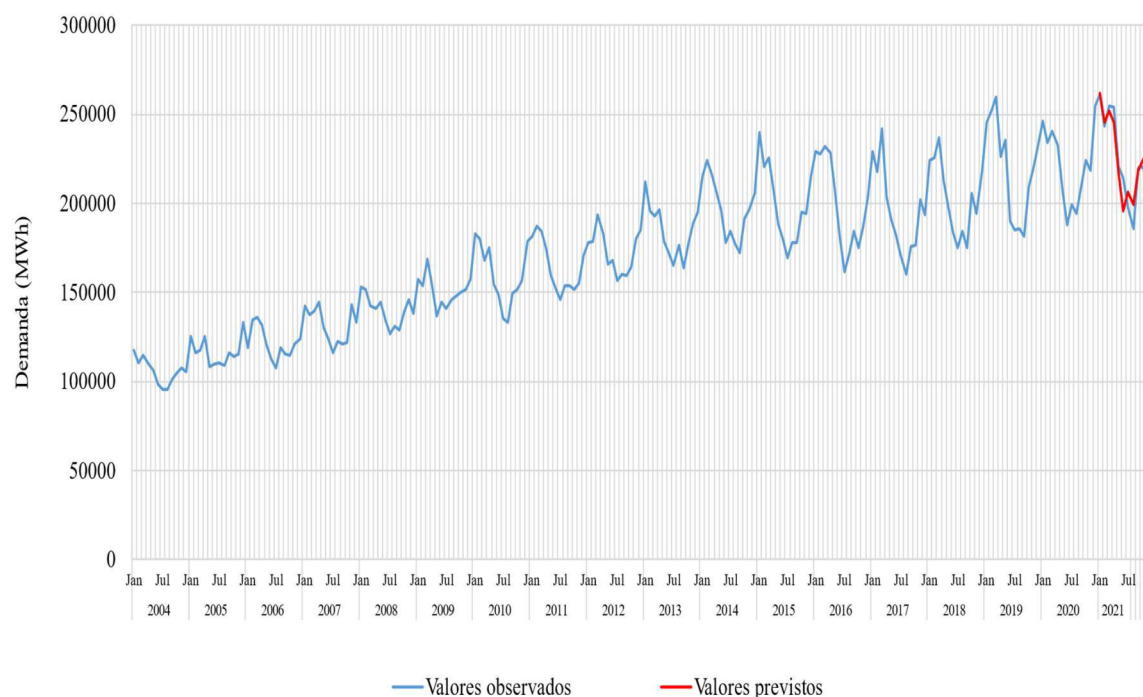


Figura 4.18- Gráfico do modelo de previsão PARIMA para demanda mensal de energia elétrica da classe residencial/ES, 2004 – 21

A partir dos modelos selecionados em cada mês, foram construídos intervalos com 95% de confiança, apresentados na Tabela 4.32, juntamente com o respectivo valor observado e valor previsto.

Tabela 4.32- Valores observados da demanda mensal de energia elétrica para a classe residencial e valores previstos e intervalo de confiança obtidos a partir de modelo PARIMA, ES, 2021

Meses	Valor observado (MWh)	Valor previsto (MWh)	IC com 95% (MWh)	
Janeiro	260909,82	262084,30	226020,30	303902,70
Fevereiro	243115,00	245548,30	213405,00	282533,00
Março	254489,90	251716,70	222320,80	284999,50
Abril	253878,90	245130,00	217642,20	276089,40
Maio	220716,80	216984,90	185300,50	254087,00
Junho	214126,50	195956,50	181463,20	211607,30
Julho	197467,90	206659,70	182201,90	234400,50
Agosto	185968,50	199068,40	167905,90	236014,30
Setembro	219803,90	218610,10	195760,40	244126,90
Outubro	220852,00	223888,10	192809,80	259975,70
Novembro	211945,40	228356,70	202152,60	257957,40
Dezembro	228721,10	253880,10	207363,30	310831,80

Fonte: Autora

A Tabela 4.32 mostra que todos os valores observados da demanda energética se encontram dentro do intervalo de confiança de 95% estabelecido de forma individual para cada subsérie e que os valores previstos subestimaram a demanda nos meses de março, abril, maio, junho e setembro, enquanto nos demais meses a demanda foi superestimada.

Esta modelagem PARIMA também foi executada com a utilização da função `auto.arima` do pacote *forecast* do *software* R. Os modelos encontrados por esta função para cada subsérie, juntamente com as medidas da capacidade preditiva encontram-se na Tabela 4.33, enquanto na Tabela 4.34 são apresentados os valores observados e previstos da demanda e o intervalo de confiança correspondente.

Tabela 4.33- Capacidade preditiva do Modelo PARIMA da demanda mensal de energia elétrica para a classe residencial, gerado a partir da função auto.arima, e respectivas medidas de acurácia, ES/2021

Meses	Modelos indicado	MAE (MWh)	MAPE (%)	RMSE (MWh)
Janeiro	ARIMA(0,1,1)	1290,31	0,4945	1290,31
Fevereiro	ARIMA(0,1,0)	2433,26	1,0008	2433,26
Março	ARIMA(2,1,0)	2943,04	1,1564	2943,04
Abril	ARIMA(0,1,1)	8857,96	3,4890	8857,96
Maio	ARIMA(0,1,0)	3731,90	1,6908	3731,90
Junho	ARIMA(0,2,1)	24037,30	11,2257	24037,30
Julho	ARIMA(0,1,0)	10809,30	5,4739	10809,30
Agosto	ARIMA(0,1,0)	17398,13	9,3554	17398,13
Setembro	ARIMA(0,1,0)	1194,24	0,5433	1194,24
Outubro	ARIMA(0,1,0)	3036,10	1,3747	3036,10
Novembro	ARIMA(0,1,0)	16411,28	7,7431	16411,28
Dezembro	ARIMA(0,1,0)	26075,10	11,4003	26075,10
Modelagem PARIMA		9851,49	4,5790	13120,75

Fonte: Autora

Tabela 4.34- Valores observados da demanda mensal de energia elétrica para a classe residencial e valores previstos e intervalo de confiança obtidos a partir de modelo PARIMA gerado a partir da função auto.arima, ES, 2021

Meses	Valor observado	Valor previsto	IC com 95%	
	(MWh)	(MWh)	(MWh)	
Janeiro	260909,82	262200,10	227187,70	302608,40
Fevereiro	243115,00	245548,30	213405,00	282533,00
Março	254489,90	251546,90	218952,10	288993,90
Abril	253878,90	245020,90	217630,90	278857,96
Maio	220716,80	216984,90	185300,50	254087,00
Junho	214126,50	190089,20	176729,50	204458,80
Julho	197467,90	208277,2	184019,40	235732,70
Agosto	185968,50	203366,60	176514,60	234303,50
Setembro	219803,90	218609,70	197152,80	242401,70
Outubro	220852,00	223888,10	192809,80	259975,70
Novembro	211945,40	228356,70	202152,60	257957,40
Dezembro	228721,10	254796,20	210825,10	307938,20

Fonte: Autora

A Tabela 4.33 mostra o mês que apresentou uma maior variabilidade dos valores previstos em relação aos observados foi o mês de dezembro. Além disso, o MAPE, também em 50% dos casos, encontrou resultados inferior a 2,5%

Todos os resultados obtidos para o grau de integração pelo teste KPSS concordaram com os indicados função autoarima, que apresentou grau de integração 2 para o mês de junho e grau de integração 1 para os outros meses. Os meses de janeiro, fevereiro, abril e maio encontraram grau de integração 1 de acordo com a função autoarima que concorda com os testes ADF, KPSS e PP e com os resultados da Tabela 4.31

Nesta modelagem os valores previstos e os valores observados da demanda energia elétrica encontram-se dentro do intervalo de confiança de 95%, estabelecido de forma individual para cada subsérie. De acordo com as tabelas 4.2.26 e 4.2.28, o MAPE em 50% dos casos encontrou resultados inferiores a 2,5%

Comparando os modelos indicados pelas duas metodologias verifica-se que o modelo PARIMA gerado pela aplicação da metodologia em etapas apresentou melhor capacidade preditiva quando comparado com o modelo PARIMA gerado pela função auto.arima, obtendo uma diferença de aproximadamente 0,5% para o MAPE, 1556,97 MWh para o RMSE e de 1091,20 MWh para o MAE.

4.3. Modelagem SARIMA

A modelagem SARIMA foi aplicada à mesma série formada pelos dados da demanda mensal residencial de energia elétrica no ES. A Tabela 4.35 apresenta os resultados dos testes de estacionariedade, aplicado aos dados de ajuste com a transformação logarítmica, indicando a ordem de integração, ou seja, a quantidade de diferenciações necessárias para serem encontradas evidências de que a série se torna estacionária. Nota-se que os dados da demanda podem ser considerados estacionários em primeira diferença na parte não sazonal da série para modelo sem tendência, enquanto na parte sazonal a série é estacionária em nível.

Tabela 4.35- Resultado do teste de estacionariedade da demanda mensal de energia elétrica para a classe consumidora residencial/ES, 2004 – 20

Parte não sazonal			Parte sazonal	
testes	Defasagem máxima 24		testes	Defasagem máxima 24
	Constante	Tendência		
ADF	1	0	CH	0
KPSS	1	1	HEGY	0
PP	1	0	OCSB	0

Fonte: Autora

As figuras 4.19 e 4.20 apresentam a FAC e FACP da parte não sazonal e da parte sazonal, respectivamente. Nestes gráficos investigou-se a ordem da componente de média móvel (MA) e da componente autoregressiva (AR), que foram construídos levando em consideração as respectivas ordens de integração sazonal e não sazonal.

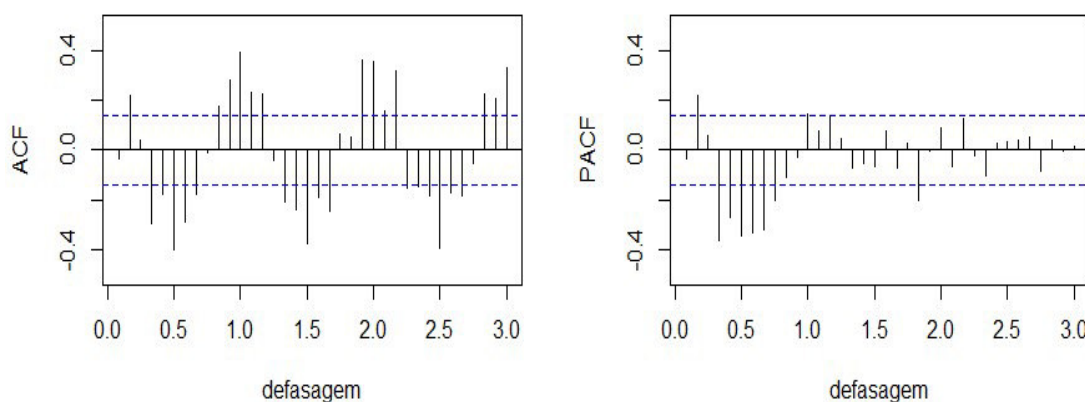


Figura 4.19- Gráficos da FAC e FACP com grau de integração 1, para a parte não sazonal

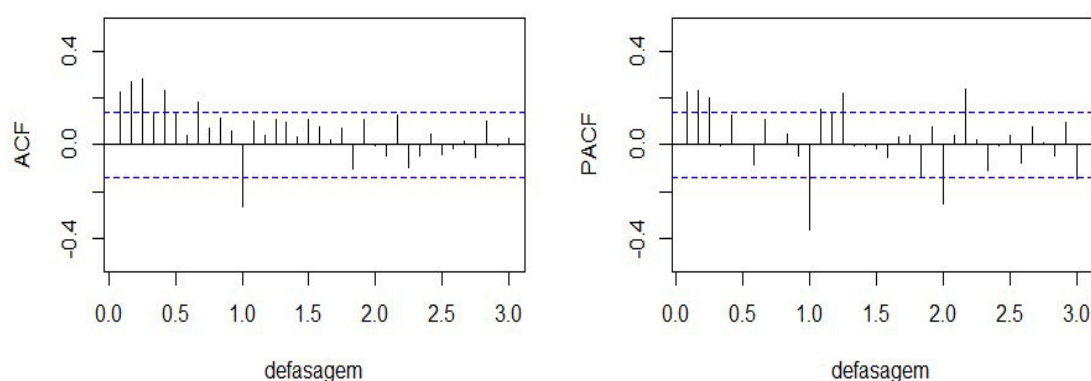


Figura 4.20- Gráficos da FAC e FACP da série em nível para a parte sazonal

Como a interpretação dos correlogramas não foi conclusiva, o que pode acontecer em função da subjetividade da interpretação gráfica, optou-se por estimar os coeficientes dos 256 modelos candidatos, determinados fazendo-se variar p , q , P e Q entre 0 e 3. Conforme apresentado na metodologia, foi avaliada a parcimônia de cada modelo através dos critérios de informação AIC, AICc e BIC, selecionando os 13 melhores para compor o conjunto de modelos candidatos, que apresentaram menores resultados para os referidos critérios. A Tabela 4.36 apresenta os modelos selecionados de acordo com AIC, AICc e BIC.

Tabela 4.36- Resultado do teste de estacionariedade da demanda mensal de energia elétrica para a classe consumidora residencial/ES, 2004 – 21

AIC	AICc	BIC
SARIMA(2,1,3)(1,0,1) ₁₂	SARIMA(2,1,3)(1,0,1) ₁₂	SARIMA(2,1,3)(1,0,1) ₁₂
SARIMA(2,1,3)(3,0,2) ₁₂	SARIMA(2,1,3)(3,0,2) ₁₂	SARIMA(2,1,3)(3,0,2) ₁₂
SARIMA(2,1,3)(1,0,0) ₁₂	SARIMA(2,1,3)(1,0,0) ₁₂	SARIMA(2,1,3)(1,0,0) ₁₂
SARIMA(3,1,3)(1,0,3) ₁₂	SARIMA(3,1,3)(1,0,3) ₁₂	SARIMA(3,1,3)(1,0,3) ₁₂
SARIMA(2,1,3)(1,0,2) ₁₂	SARIMA(2,1,3)(1,0,2) ₁₂	SARIMA(0,1,1)(1,0,2) ₁₂
SARIMA(2,1,3)(1,0,3) ₁₂	SARIMA(2,1,3)(1,0,3) ₁₂	SARIMA(2,1,3)(1,0,2) ₁₂
SARIMA(2,1,3)(2,0,3) ₁₂	SARIMA(2,1,3)(2,0,3) ₁₂	SARIMA(0,1,1)(2,0,1) ₁₂
SARIMA(1,1,3)(1,0,3) ₁₂	SARIMA(1,1,3)(1,0,3) ₁₂	SARIMA(1,1,1)(1,0,1) ₁₂
SARIMA(2,1,3)(3,0,1) ₁₂	SARIMA(3,1,1)(1,0,3) ₁₂	SARIMA(0,1,2)(1,0,1) ₁₂
SARIMA(3,1,1)(1,0,3) ₁₂	SARIMA(2,1,3)(3,0,1) ₁₂	SARIMA(2,1,1)(1,0,1) ₁₂
SARIMA(2,1,1)(1,0,3) ₁₂	SARIMA(2,1,1)(1,0,3) ₁₂	SARIMA(0,1,3)(1,0,1) ₁₂
SARIMA(3,1,2)(1,0,3) ₁₂	SARIMA(0,1,3)(1,0,3) ₁₂	SARIMA(0,1,1)(1,0,3) ₁₂
SARIMA(0,1,3)(1,0,3) ₁₂	SARIMA(1,1,3)(1,0,1) ₁₂	SARIMA(3,1,1)(1,0,1) ₁₂

Fonte: Autora

A Tabela 4.36 mostra que os critérios AIC e AICc apresentaram os mesmos resultados em cerca de 92% dos modelos classificados como mais parcimoniosos, apontando apenas diferença em apenas um modelo,entretanto o BIC concorda com estes resultados em apenas 46% dos casos. Assim, foram considerados 21 modelos para avaliação subsequente.

A Tabela 4.37 apresenta os valores dos critérios de informação bem como o resultado da averiguação da presença de autocorrelação, normalidade e 73omoscedasticidade nos resíduos dos modelos selecionados.

Tabela 4.37- Modelos candidatos escolhidos pela parcimoniaalidade para a demanda mensal de energia elétrica para a classe consumidora residencial, ES/2004-21

Modelos	Critérios de informação			Análise de resíduos		
	AIC	AICc	BIC	JB (p-valor)	LB (p-valor)	ARCH (p-valor)
SARIMA(2,1,3)(1,0,1) ₁₂	-654,33	-653,40	-624,51	0,2961	0,6621	0,7954
SARIMA(2,1,3)(3,0,2) ₁₂	-650,10	-648,46	-610,34	0,3330	0,4934	0,5038
SARIMA(2,1,3)(1,0,0) ₁₂	-644,08	-643,34	-617,57	0,2688	0,4282	0,9804
SARIMA(3,1,3)(1,0,3) ₁₂	-637,45	-635,81	-597,69	0,8585	0,3234	0,4973
SARIMA(2,1,3)(1,0,2) ₁₂	-636,25	-635,11	-603,12	0,7320	0,1301	0,6814
SARIMA(2,1,3)(1,0,3) ₁₂	-635,63	-634,25	-599,18	0,7440	0,1450	0,6674
SARIMA(2,1,3)(2,0,3) ₁₂	-634,67	-633,03	-594,91	0,5895	0,1620	0,4655
SARIMA(1,1,3)(1,0,3) ₁₂	-628,00	-626,85	-594,86	0,8231	0,3877	0,5372
SARIMA(2,1,3)(3,0,1) ₁₂	-627,82	-626,44	-591,38	0,8901	0,2026	0,6472
SARIMA(3,1,1)(1,0,3) ₁₂	-627,81	-626,66	-594,68	0,8776	0,4764	0,6554
SARIMA(2,1,1)(1,0,3) ₁₂	-627,26	-626,33	-597,44	0,8409	0,2074	0,5552
SARIMA(3,1,2)(1,0,3) ₁₂	-627,15	-625,77	-590,70	0,8935	0,3044	0,6069
SARIMA(0,1,3)(1,0,3) ₁₂	-626,98	-626,05	-597,16	0,8892	0,2281	0,6077
SARIMA(1,1,3)(1,0,1) ₁₂	-626,64	-625,89	-600,13	0,8956	0,1548	0,6183
SARIMA(0,1,1)(1,0,2) ₁₂	-623,18	-622,75	-603,30	0,9948	0,0268	0,8947
SARIMA(0,1,1)(2,0,1) ₁₂	-622,96	-622,53	-603,08	0,9954	0,0285	0,9024
SARIMA(1,1,1)(1,0,1) ₁₂	-622,33	-621,91	-602,45	0,9963	0,0302	0,8874
SARIMA(0,1,2)(1,0,1) ₁₂	-622,32	-621,89	-602,44	0,9953	0,0304	0,8980
SARIMA(2,1,1)(1,0,1) ₁₂	-625,28	-624,70	-602,08	0,9047	0,0574	0,6030
SARIMA(0,1,3)(1,0,1) ₁₂	-624,70	-624,13	-601,51	0,9597	0,0782	0,7405
SARIMA(0,1,1)(1,0,3) ₁₂	-623,81	-623,24	600,62	0,9779	0,0343	0,8539

De acordo com os três critérios de informação, conforme Tabela 4.37, o modelo SARIMA(2,1,1)(1,0,3)₁₂ apresentou melhores resultados, sendo mais parcimonioso. Ainda nesta tabela, considerando nível de significância 0,05, verifica-se que os modelos SARIMA(0,1,1)(1,0,2)₁₂, SARIMA (0,1,1)(2,0,1)₁₂, SARIMA (1,1,1)(1,0,1)₁₂, SARIMA (0,1,2)(1,0,1)₁₂ e SARIMA(0,1,1)(1,0,3)₁₂, selecionados pelo critério de informação BIC, apresentaram evidências de autocorrelação nos resíduos, uma vez que a hipótese nula foi rejeitada, razão pela qual foram eliminados da análise subsequente. Para os testes de normalidade e homoscedasticidade, as hipóteses nulas não foram rejeitadas, indicando que os resíduos de todos os modelos podem ser considerados homocedásticos e com distribuição normal.

A Tabela 4.38 apresenta as medidas de acurácia: o erro absoluto médio (MAE), o erro absoluto percentual médio (MAPE), raiz do erro quadrático médio (RMSE), utilizadas para avaliar a capacidade preditiva.

Tabela 4.38- Indicadores de qualidade para validação dos modelos SARIMA da classe consumidora residencial/ ES,2004-21

Modelo selecionado	MAE (MWh)	MAPE (%)	RMSE (MWh)
SARIMA(3,1,2)(1,0,3) ₁₂	8672,89	3,8954	10744,73
SARIMA(2,1,1)(1,0,3) ₁₂	9165,88	4,0817	10540,17
SARIMA(1,1,3)(1,0,3) ₁₂	9279,91	4,1318	10697,36
SARIMA(0,1,3)(1,0,3) ₁₂	9287,89	4,1497	10825,64
SARIMA(3,1,1)(1,0,3) ₁₂	9826,30	4,3587	11264,60
SARIMA(3,1,3)(1,0,3) ₁₂	10153,42	4,5087	12028,06
SARIMA(1,1,3)(1,0,1) ₁₂	10956,87	5,0050	12849,72
SARIMA(2,1,1)(1,0,1) ₁₂	11017,55	5,0367	12928,15
SARIMA(0,1,3)(1,0,1) ₁₂	11151,62	5,1054	13199,48
SARIMA(2,1,3)(3,0,1) ₁₂	11778,09	5,3534	13615,37
SARIMA(2,1,3)(1,0,0) ₁₂	11636,77	5,2983	15387,03
SARIMA(2,1,3)(1,0,1) ₁₂	13084,41	5,8381	16287,67
SARIMA(2,1,3)(2,0,3) ₁₂	13392,03	6,0328	15900,29
SARIMA(2,1,3)(1,0,3) ₁₂	13433,47	6,0448	15802,33
SARIMA(2,1,3)(3,0,2) ₁₂	13919,15	6,1829	17216,67
SARIMA(2,1,3)(1,0,2) ₁₂	14325,83	6,4424	16769,15

De acordo com a Tabela 4.38 o modelo SARIMA(3,1,2)(1,0,3)₁₂ apresenta menores

valores para os indicadores de qualidade MAE e MAPE, enquanto o modelo SARIMA(2,1,1)(1,0,3)₁₂ possui menor valor para o indicador de qualidade RMSE. Dessa maneira, tem-se dois modelos candidatos que podem ser escolhidos para previsão da demanda.

Em função dessa divergência, para a identificação do melhor modelo, ainda foram considerados também o fato dos valores observados da série estarem dentro dos intervalos de confiança, além de resultados da análise de resíduos, dos critérios de informação e do RMSE. Segundo Morettin e Toloi (2018), este indicador é o melhor por dar mais peso aos resíduos maiores.

Os dois modelos apresentaram todos os valores observados dentro do intervalo de confiança, conforme o apresentado nas tabelas 4.39 e 4.40. Observou-se que, em ambos os modelos, o valor previsto subestimou a demanda nos meses de: abril, junho, setembro e outubro.

Tabela 4.39- Valores observados da demanda mensal de energia elétrica para a classe residencial e valores previstos e intervalo de confiança obtidos a partir de modelo SARIMA(3,1,2)(1,0,3)₁₂, ES,2021

meses	Valor observado (MWh)	Valor previsto (MWh)	IC com 95% (MWh)	
Janeiro	260909,82	261191,60	238299,70	286282,70
Fevereiro	243115,00	261329,60	237535,80	287506,70
Março	254489,90	266233,70	240616,70	294577,80
Abriu	253878,90	246772,60	222360,20	273865,30
Maio	220716,80	225465,90	202730,30	250751,10
Junho	214126,50	209827,00	188210,40	233926,30
Julho	197467,90	197582,80	176953,70	220616,80
Agosto	185968,50	202396,00	180870,50	226483,40
Setembro	219803,90	204506,80	182500,40	229166,90
Outubro	220852,00	220269,60	196165,80	247335,20
Novembro	211945,40	223907,80	199136,40	251760,60
dezembro	228721,10	242016,60	214821,80	272654,20

Tabela 4.40- Valores observados da demanda mensal de energia elétrica para a classe residencial e valores previstos e intervalo de confiança obtidos a partir de modelo SARIMA(2,1,1)(1,0,3)₁₂, ES,2021

meses	Valor observado (MWh)	Valor previsto (MWh)	IC com 95% (MWh)	
Janeiro	260909,82	264107,70	241014,90	289413,10
Fevereiro	243115,00	259632,00	236019,60	285606,60
Março	254489,90	268815,30	242897,40	297498,60
Abriu	253878,90	245233,20	221002,00	272121,10
Mai	220716,80	226741,90	203819,80	252241,80
Junho	214126,50	208489,60	187029,80	232411,70
Julho	197467,90	197952,40	177233,20	221093,80
Agosto	185968,50	200570,90	179250,70	224426,90
Setembro	219803,90	205082,60	182959,00	229881,60
Outubro	220852,00	217874,90	194036,60	244641,90
Novembro	211945,40	224320,60	199439,50	252305,70
dezembro	228721,10	239203,40	212318,30	269492,80

Na análise da distribuição dos resíduos, conforme figuras 4.21 e 4.22, verifica-se que o Histograma e o Box Plot tem aparências similares, embora o Histograma dos resíduos do modelo SARIMA(2,1,1)(1,0,3)₁₂ apresente uma lacuna e o *BoxPlot* do modelo SARIMA(3,1,2)(1,0,3)₁₂ apresente um possível valor fora do padrão

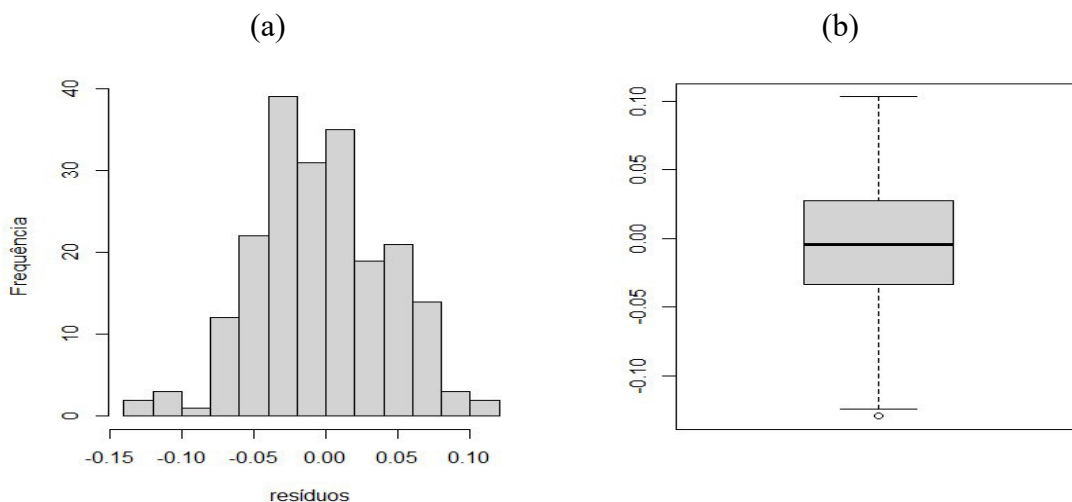


Figura 4.21- Histograma (a) e Box Plot (b) do modelo SARIMA(3,1,2)(1,0,3)₁₂ para demanda mensal de energia elétrica da classe residencial/ES, 2021

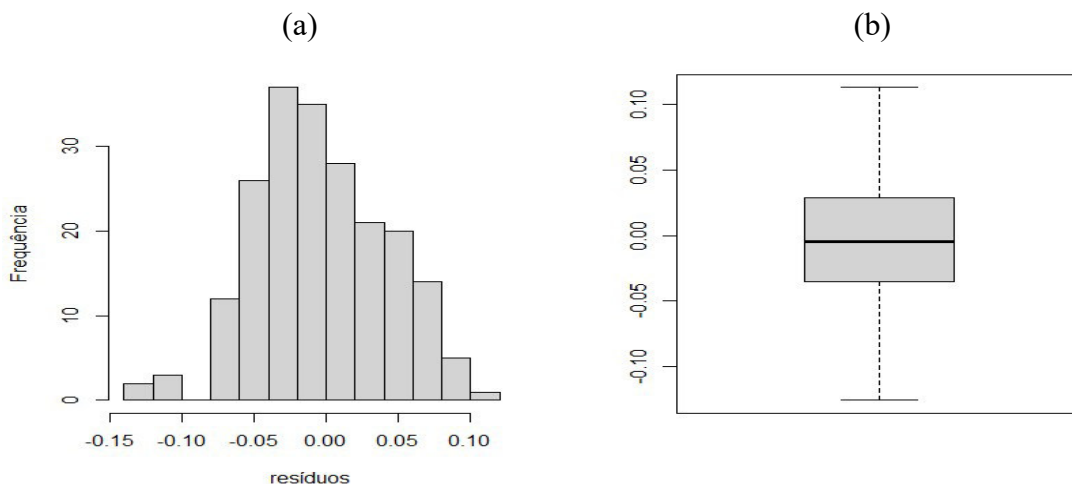


Figura 4. 22- Histograma (a) e Box Plot (b) do modelo SARIMA(2,1,1)(1,0,3)₁₂ para demanda mensal de energia elétrica da classe residencial/ES, 2021

Considerando que o modelo SARIMA(2,1,1)(1,0,3)₁₂ apresentou menores resultados para os três critérios de informação AIC, AICc e BIC, sendo mais parcimonioso, não possui nenhum valor observado fora do intervalo de confiança, obteve melhor resultado de acordo com o RMSE e a distribuição de seus resíduos é similar ao outros candidato, ele foi escolhido para prever a demanda energética para o ano de 2021.

A Figura 4.23 mostra o gráfico que apresenta o modelo SARIMA(2,1,1)(1,0,3)₁₂ que foi identificado como apresentando melhor capacidade preditiva.

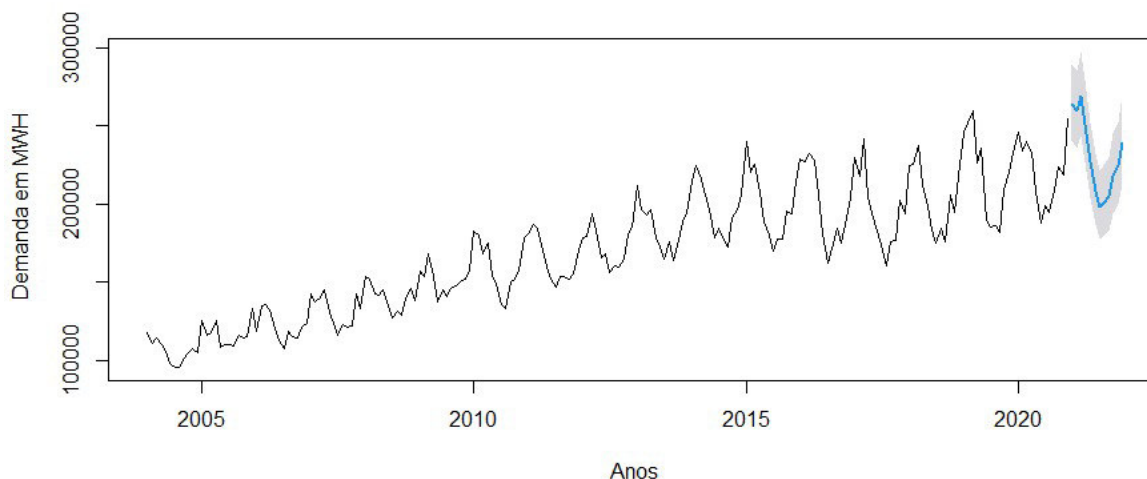


Figura 4.23- Gráfico do modelo de previsão SARIMA(2,1,1)(1,0,3)₁₂ para demanda mensal de energia elétrica da classe residencial/ES, 2004 - 21

Depois desta análise, investigou-se os modelos gerados pela função `auto.arima` do pacote *forecast* do *software R*. A Tabela 4.41 apresenta os modelos sugeridos de acordo com os resultados dos critérios AIC, AICc e BIC, em que se verifica que o modelo SARIMA(2,1,1)(1,0,3)₁₂ (AIC=-627,26; AICc=-626,33 e BIC=-597,44), apresenta melhores resultados para os critérios de informação do que estes.

Tabela 4.41- Resultados dos critérios de informação AIC, AICc e BIC para os modelos considerados como melhores pela função `auto.arima` para a demanda de energia elétrica da classe residencial/ ES,2004-21

Modelos	AIC	AICc	BIC
SARIMA(2,1,2)(1,1,1) ₁₂	-597,93	-597,31	-575,16
SARIMA(0,1,0)(0,1,0) ₁₂	-419,56	-419,54	-416,31
SARIMA(1,1,0)(1,1,0) ₁₂	-528,51	-528,39	-518,76
SARIMA(0,1,1)(0,1,1) ₁₂ *	-603,65	-603,52	-593,90
SARIMA(0,1,1)(0,1,0) ₁₂	-518,38	-518,32	-511,87
SARIMA(0,1,1)(1,1,1) ₁₂	-602,33	-602,12	-589,32
SARIMA(0,1,1)(0,1,2) ₁₂	-602,56	-602,35	-589,55
SARIMA(0,1,1)(1,1,0) ₁₂	-564,29	-564,16	-554,53
SARIMA(0,1,0)(0,1,1) ₁₂	-518,16	-518,09	-511,65
SARIMA(1,1,1)(0,1,1) ₁₂	-602,10	-601,89	-589,09
SARIMA(0,1,2)(0,1,1) ₁₂	-602,00	-601,78	-588,99
SARIMA(1,1,0)(0,1,1) ₁₂	-578,61	-578,49	-568,86
SARIMA(1,1,2)(0,1,1) ₁₂	-600,21	-599,88	-583,95

Obs * modelo considerado como mais pacimonioso pelo `autoarima`.

No modelo SARIMA(0,1,1)(0,1,1)₁₂, considerado como apresentando melhor ajuste por ser mais parcimonioso pela função auto.arima, foi realizada a análise de resíduos para verificação dos pressupostos teóricos. O modelo cumpre os requisitos de normalidade (valor $p = 0,6266$ no teste Jarque-Bera), homocedasticidade (valor $p = 0,8931$ no teste ARCH) e independência (valor $p = 0,2195$ no teste Ljung-Box). Os resultados para os indicadores da capacidade preditiva MAE, MAPE e RMSE assumiram os valores 12310,89MWh, 5,5997% e 14316,30MWh, respectivamente, piores dos já encontrados pelo outro protocolo.

A Tabela 4.42 apresenta os valores observados e previstos, com os respectivos intervalos construídos com 95% de confiança, onde se constata que o valor observado para a demanda mensal de energia residencial no mês de agosto encontra-se fora do intervalo de confiança, possuindo um valor observado menor que o limite inferior do intervalo de confiança.

Tabela 4.42- Valores observados da demanda mensal de energia elétrica para a classe residencial e valores previstos e intervalo de confiança obtidos a partir de modelo SARIMA(0,1,1)(0,1,1)₁₂, ES, 2021

Meses	Valor observado (MWh)	Valor previsto (MWh)	IC com 95% (MWh)	
Janeiro	260909,82	265250,60	241503,40	291332,90
Fevereiro	243115,00	260567,40	236580,80	286986,00
Março	254489,90	268787,40	243385,50	296840,50
Abril	253878,90	249404,80	225241,40	276160,50
Maiο	220716,80	233178,80	210048,00	258856,80
Junho	214126,50	211618,60	190150,00	235511,00
Julho	197467,90	206336,20	184950,60	230194,60
Agosto	185968,50	208028,80	186022,20	232638,80
Setembro	219803,90	212055,50	189179,30	237697,90
Outubro	220852,00	228112,20	203037,40	256283,80
Novembro	211945,40	234066,00	207868,30	263565,40
Dezembro	228721,10	252859,20	224062,60	285356,70

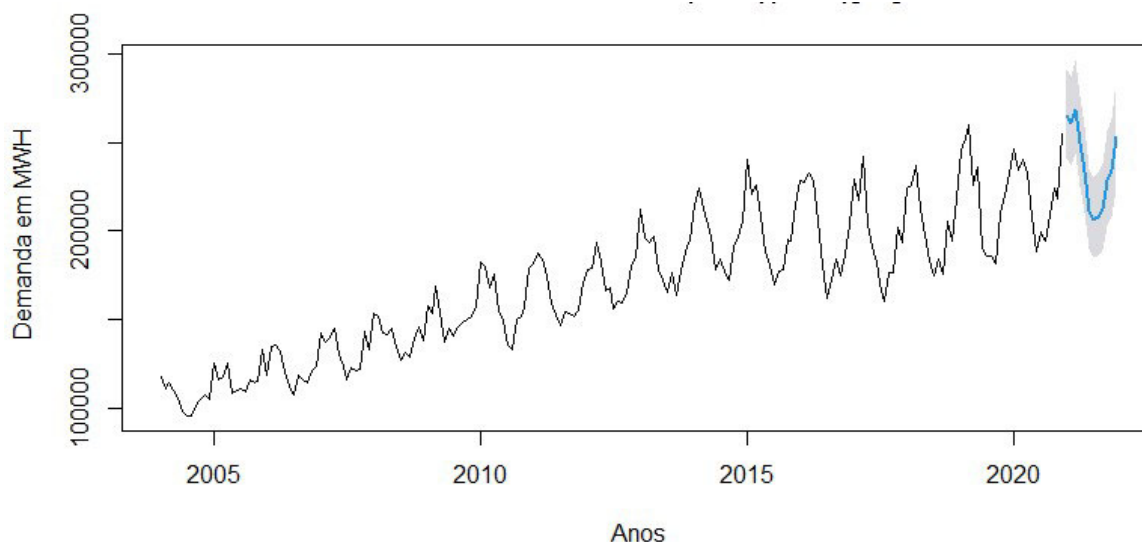


Figura 4.24- Gráfico de previsão do modelo SARIMA(0,1,1)(0,1,1)₁₂ para a demanda mensal de energia elétrica da classe residencial, ES/ 2004-21

Verifica-se que o modelo SARIMA gerado pela seleção do melhor modelo candidato, SARIMA(2,1,1)(1,0,3)₁₂, da mesma forma que ocorreu na modelagem PARIMA, encontrou melhor capacidade preditiva do que o modelo SARIMA gerado pela função `auto.arima`, SARIMA(0,1,1)(0,1,1)₁₂.

4.4. Considerações sobre os resultados das modelagens PARIMA e SARIMA

De acordo com a Tabela 4.43, os resultados dos indicadores MAE e MAPE para a modelagem PARIMA superaram os obtidos pela modelagem SARIMA, porém, o modelo SARIMA obteve melhor resultado para o RMSE. Tanto o modelo SARIMA quanto o modelo PARIMA não mostraram valores fora do intervalo de confiança gerado pelos modelos.

Tabela 4.43- Capacidade preditiva obtidas nas modelagens PARIMA e SARIMA

Medida	Modelagem	
	PARIMA	SARIMA
MAE (MWh)	8760,29	9165,88
MAPE (%)	4,0435	4,0817
RMSE (MWh)	13120,75	10540,17

As figuras 4.25 e 4.26 mostram os gráficos que apresentam os valores observados, previstos e os respectivos intervalos de confiança dos modelos PARIMA e SARIMA, respectivamente.

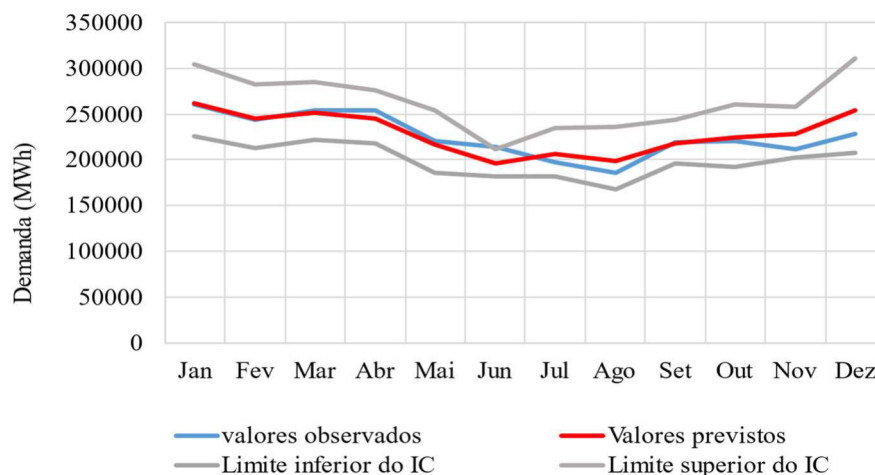


Figura 4.25- Gráfico da previsão para a demanda residencial ES/ 2021 gerado pelo modelo PARIMA

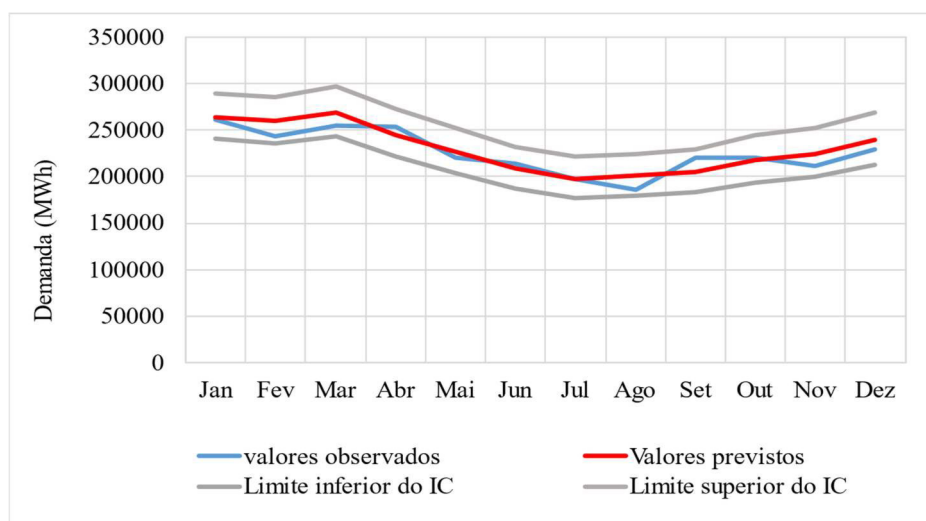


Figura 4.26- Gráfico da previsão para a demanda residencial ES/ 2021 gerado pelo modelo SARIMA

Analisando o resultado apresentado pelas duas abordagens da metodologia de Box-Jenkins, tem-se que os modelos SARIMA e PARIMA apresentaram resultados próximos. Porém, nos primeiros meses a modelagem PARIMA parece ter obtido melhores previsões, enquanto que a modelagem SARIMA, nos últimos meses. Além disso, o modelo SARIMA(2,1,1)(1,0,3)₁₂ apresentou erros de menor intensidade, o que levou a um menor RMSE.

4.5. Considerações sobre o desempenho das técnicas utilizadas nas modelagens PARIMA e SARIMA

Para identificar a sazonalidade, tanto os métodos analíticos quanto os métodos gráficos se mostraram eficientes, porém, os métodos gráficos por possibilitarem a visualização dos dados, facilitam a compreensão dos resultados.

Na etapa de análise de estacionariedade, na abordagem PARIMA, foram empregados os testes ADF, KPSS e PP. Notou-se que o KPSS concordou com os resultados obtidos pela função `auto.arima`, no entanto, os testes ADF e PP apresentaram valores discrepantes para o grau de integração, chegando a sugerir grau de integração cinco. De modo geral o teste ADF foi o que mais oscilou nos resultados. Na abordagem SARIMA foram empregados os testes ADF, KPSS e PP na parte não sazonal e os testes Ch, Hegy e OCSB na parte sazonal. Seus desempenhos para a parte sazonal foram considerados satisfatórios por haver concordância de resultados, o mesmo não acontecendo em relação à parte não sazonal.

Ainda na aplicação dos testes que avaliam a estacionariedade, os resultados do teste ADF apresentaram melhores resultados quando adotado defasagem máxima 2 (critério de Hyndman e Athanasopoulos) quando comparados com os resultados para defasagem máxima 7 (critério de Schwert).

A análise dos gráficos das funções de autocorrelação e autocorrelação parcial, FAC e FACP, se mostrou inconclusiva, sendo necessário analisar uma grande quantidade de modelos a partir da determinação de valor máximo e mínimo para p , q , P e Q .

Na análise dos resultados encontrados pelos critérios de informação, notou-se que os critérios AIC e AICc apresentaram resultados equivalentes na identificação dos modelos mais parcimoniosos, enquanto o critério BIC não identificou os modelos SARIMA(3,1,2)(1,0,3)₁₂ e SARIMA(2,1,1)(1,0,3)₁₂ que foram considerados como os modelos com melhor capacidade preditiva. Vale ressaltar que o modelo que apresentou melhor capacidade preditiva foi identificado pelos critérios AIC e AICc.

De acordo com a análise da capacidade preditiva, observou-se que o MAE e MAPE, apresentaram melhores valores para o modelo SARIMA(3,1,2)(1,0,3)₁₂ e o RMSE para o modelo SARIMA(2,1,1)(1,0,3)₁₂. Quando realizada a análise de resíduos para desempate, o modelo classificado pelo RMSE apresentou-se mais homogêneo, sendo considerado como o critério que identificou o modelo com melhor capacidade preditiva.

Portanto, o modelo considerado como apresentando melhor desempenho foi encontrado por meio da modelagem SARIMA, seguindo a modelagem passo a passo. Em

algumas das etapas deste protocolo, algumas técnicas, apresentadas no Quadro 4.1, apresentaram melhor desempenho.

Quadro 4.1 – Técnicas recomendadas

Etapas da análise	Técnicas
Análise exploratória de dados	Gráficos em linhas e das subséries
Estacionariedade	Parte não sazonal- KPSS
Modelos candidatos	Determinar valor máximo e mínimo para p e q, analisando todos os modelos gerados e selecionar os mais parcimoniosos
Parcimonialidade	AIC e AICc
Capacidade preditiva	RMSE e análise da distribuição dos resíduos.

Fonte: Autora

5. CONSIDERAÇÕES FINAIS

O presente estudo teve como objetivo modelar a demanda mensal de energia elétrica para a classe residencial do estado do Espírito Santo, utilizando as abordagens PARIMA e SARIMA oriundas da metodologia de Box-Jenkins. A motivação para a realização deste estudo reside na intenção de oferecer apoio ao governo do estado do Espírito Santo em programas como o GERAR e em suas tomadas de decisão. Isto é particularmente relevante porque a previsão da demanda de energia elétrica desempenha um papel crucial na asseguarção de um fornecimento energético confiável, sustentável e eficiente, além de impulsionar o crescimento econômico e a sensibilização para questões ambientais.

Para isso, foi realizada uma análise exploratória na qual foram identificadas algumas propriedades dos dados que compõem a série temporal analisada, como a presença de sazonalidade. Após a análise exploratória, a série foi modelada aplicando as abordagens PARIMA e SARIMA por meio de diversos protocolos. Os resultados mostraram que, de acordo com o MAE e MAPE, o modelo PARIMA obteve melhor desempenho na realização da previsão. No entanto, o RMSE identificou o modelo SARIMA como o de melhor capacidade preditiva.

Os resultados encontrados neste trabalho podem ser contrastados com os obtidos por Pereira (2020). A autora realizou a modelagem da demanda de energia elétrica para as classes consumidoras residencial, comercial e industrial, do estado do Rio Grande do Sul, utilizando as metodologias SARIMA e SARIMA com a inserção de variáveis exógenas (SARIMAX). Para avaliar a parcimônia dos modelos, foram utilizados o critério BIC e os indicadores MAE, MAPE, RMSE e a estatística U de Theil para avaliar a capacidade preditiva dos modelos.

Com relação aos resultados obtidos, verificou-se que, em geral, quanto menos defasagens são inseridas na variável externa na modelagem, mais parcimoniosos são os modelos desenvolvidos. Para a classe residencial, o modelo escolhido não foi o mais parcimonioso e considerou apenas uma variável externa, a tarifa com 1 defasagem.

Uma hipótese a ser avaliada em trabalhos futuros é a discordância de resultados entre o teste ADF e outros testes para verificação da estacionariedade. Outra discordância ocorreu nos resultados do critério de informação BIC, que divergiu dos resultados dos outros critérios na modelagem SARIMA. Além disso, uma limitação identificada foi a incapacidade de estimar os 256 modelos candidatos, mesmo utilizando os métodos de otimização disponibilizados pela função “arima” do pacote “forecast” do software R.

REFERÊNCIAS

- AKAIKE, H. Maximum likelihood identification of Gaussian autoregressive moving average models. **Biometrika**, v. 60, n. 2, p. 255-265, 1973.
- ANDOH, P. Y. A.; SEKYERE, C.K.K.; MENSAH, L.D.; DZEBRE, D. E. K. Forecasting electricity demand in Ghana with the SARIMA model. **Journal of Applied Engineering and Technological Science (JAETS)**, v. 3, n.1, p.1-9, 2021.
- ASPE (Agência de Serviços Públicos de Energia do Estado do Espírito Santo). **Balanço Energético do Estado do Espírito Santo – Ano Base 2014**. Vitória (ES): ASPE, 2022a.
- Associação Brasileira de Energia Solar - ABSOLAR**. Disponível em: < <https://www.absolar.org.br/noticias> >. Acesso em: março de 2023.
- BORGES, F. Q. Crise de energia elétrica no Brasil-uma breve reflexão sobre a dinâmica de suas origens e resultados. **RECIMA21-Revista Científica Multidisciplinar-ISSN 2675-6218**, v. 2, n. 10, p. e210809-e210809, 2021.
- BOZKURT, Ö. Ö.; BIRICIK, G.; TAYSI, Z. C. Artificial neural network and SARIMA based models for power load forecasting in Turkish electricity market. **PloS one**, v. 12, n. 4, p. e0175915, 2017
- BUENO, R. L. S. **Econometria de séries temporais**. 2 ed, São Paulo: Editora Cengage Learning, 2011, 124p.
- CANOVA, F.; HANSEN, B. E. Are seasonal patterns constant over time? A test for seasonal stability. **Journal of Business & Economic Statistics**, v. 13, n. 3, p. 237-252, 1995.
- CARO, E.; JUAN, J.; CARA, J. Periodically correlated models for short-term electricity load forecasting. **Applied Mathematics and Computation**, v. 364, p. 124642, 2020.
- COSER, M. C.; MENDONÇA, A.S.F. Modelagem de séries de vazões sazonais apresentando dependência de longo termo. **SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS**, v. 19, 2009.
- CRYER, J. D.; CHAN, K. **Time Series Analysis Whit Applicatons in R**. 2º ed, New York: Spring, 2008.
- CRUZ, A. *et al.* The effect of wind generation and weekday on Spanish electricity spot price forecasting. **Electric Power Systems Research**, v. 81, n. 10, p. 1924-1935, 2011.
- DARBELLAY, G. A.; SLAMA, M. Forecasting the short-term demand for electricity: Do neural networks stand a better chance?. **International Journal of Forecasting**, v. 16, n. 1, p. 71-83, 2000.
- DICKEY, D. A.; FULLER, W.A. Likelihood ratio statistics for autoregressive time series with a unit root. **Econometrica: journal of the Econometric Society**, p. 1057-1072, 1981.

EPE, Empresa de Pesquisa Energética, Consumo Mensal de Energia Elétrica por classe (regiões e subsistemas), Rio de Janeiro, Brasil, 2022 Disponível em:<<https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/Consumo-mensal-de-energia-eletrica-por-classe-regioes-e-subsistemas>>. Acesso em: março, de 2022

ERDOGDU, E. Electricity demand analysis using cointegration and ARIMA modelling: A case study of Turkey. **Energy policy**, v. 35, n. 2, p. 1129-1146, 2007.

FERREIRA, R. V. **Previsão de demanda: um estudo de caso para o sistema interligado nacional**. 2006.

FREITAS. Aprovado projeto de lei do programa de geração de energias renováveis do Espírito Santo. **Governo ES**. Disponível em < <https://www.es.gov.br/Noticia/aprovado-projeto-de-lei-do-programa-de-geracao-de-energias-renovaveis-do-espírito-santo>> acesso em: março de 2022

GRAVES, S.; GRAVES, M. S. Package ‘FinTS’. 2009.

GUJARATI, D.N.; PORTER, D. C. **Econometria Básica**. 5ªed, Porto Alegre: Editora AMG, 2011. 920p.

HILL, R. C.; GRIFFITHS, W. E.; JUDGE, G. G. **Econometria**. 2º ed. Sao Paulo: Editora Saraiva, 2003.

HOTHORN, T *et al.* lmtest: **Testing linear** HAMNER, B.; FRASCO, M. **Metrics: Evaluation Metrics for Machine Learning**. R package version 0.1.4. <https://CRAN.R-project.org/package=Metrics>.2018

HYNDMAN, R. *et al.* **forecast: Forecasting functions for time series and linear models**.R packageversion 8.16, 2022, <https://pkg.robjhyndman.com/forecast/>

HYNDMAN, R. J.; KHANDAKAR, Y. **Automatic time series forecasting: the forecast package for R**. Journal of statistical software, v. 27, p. 1-22, 2008.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: principles and practice**, 3º ed, Melbourne, Austrália: OTexts, 2018. Disponível em: <<https://otexts.com/fpp3/>>. Acesso em: jan. 2022.

HYLLEBERG, S. *et al.* Seasonal integration and cointegration. **Journal of econometrics**, v. 44, n. 1-2, p. 215-238, 1990.

JARQUE, C. M.; BERA, A. K. A Test for Normality of Observations and Regression Residuals. **International Statistical Review**, 55(2):163-172, 1986.

KWIATKOWSKI, D. *et al.* Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?. **Journal of econometrics**, v. 54, n. 1-3, p. 159-178, 1992.

LANA, L. T. C. *et al.* Energia solar fotovoltaica: revisão bibliográfica. **Engenharias On-line**, v. 1, n. 2, p. 21-33, 2015.

LEVENE, H. *et al.* Contributions to probability and statistics. **Essays in honor of Harold Hotelling**, v. 278, p. 292, 1960.

LJUNG, G. M.; BOX, G. E. On a measure of lack of fit in time series models. **Biometrika**, v. 65, n. 2, p. 297-303, 1978.

LUZ, H. P. B. **Crescimento econômico e consumo de energia: uma abordagem econométrica com séries temporais para a economia brasileira no período de 1970-2000**. 2003. Dissertação de Mestrado. Universidade Federal de Pernambuco.

MARTINS, L. O. S. *et al.* Dynamics of Industrial Electricity Demand in the State of Bahia (Brazil): Evolution of Price and Income and COVID-19 Implications. **International Journal of Energy Economics and Policy**, v. 13, n. 2, p. 1, 2023.

MARQUES, F. M. R. Perspectivas para a Energia Solar no Brasil. **Revista Business School São Paulo**, v. 1, p. 10-13, 2014.

MATTOS, V. L. D. *et al.* Modeling the commercial electricity demand in Santa Catarina, using the Box-Jenkins methodology. **International Journal of Development Research**, v. 11, n. 06, p. 48190-48197, 2021.

MEYER, D. *et al.* e1071: **Misc Functions of the Department of Statistics, Probability Theory Group** –(Formerly:E1071), TU Wien. R package version 1.7-7. 2017. <https://CRAN.R-project.org/package=e1071>

MORETTIN, P. A.; TOLOI, C.M. C. **Análise de Séries Temporais Modelos lineares univariados**, 3º ed. São Paulo: Blücher, 474 p., 2018. ISBN: 978-85-212-1352-9

MACMACKIN, N; MILLER, L; CARRIVEAU, R. Modeling and disaggregating hourly effects of weather on sectoral electricity demand. **Energy**, v. 188, p. 115956, 2019

NUNES, G. S *et al.* Avaliação da capacidade preditiva de modelos ARIMA e VAR-VEC: o caso da demanda por energia elétrica no Rio Grande do Sul. **Exacta**, v.20, p. 307-325, 2022.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, v. 52, n. 3/4, p. 591-611, 1965.

SIGAUKE, C. Forecasting medium-term electricity demand in a South African electric power supply system. **Journal of Energy in Southern Africa**, v. 28, n. 4, p. 54-67, 2017.

PHILLIPS, P. C.; PERRON, P. Testing for a unit root in time series regression. **Biometrika**, v. 75, n. 2, p. 335-346, 1988.

OSBORN, D. R. *et al.* Seasonality and the Order of Integration for Consumption, **EconPapers**, v. 50, n. 4, p.361 - 377, 1988.

PAGOTTO. Aprovado projeto de lei do programa de geração de energias renováveis do Espírito Santo. **Governo ES**. Disponível em < <https://www.es.gov.br/Noticia/aprovado-projeto-de-lei-do-programa-de-geracao-de-energias-renovaveis-do-espírito-santo> > acesso em: março de 2022

PEREIRA, T. P. **O uso dos métodos SARIMA e SARIMAX na modelagem da demanda mensal de energia elétrica no estado do Rio Grande do Sul**. 2021. Dissertação de Mestrado, RS: Universidade Federal do Rio Grande (FURG).

PFÄFF, Bernhard. **Analysis of integrated and cointegrated time series with R**. Springer Science & Business Media, 2008.

POLPRASERT, J.; NGUYÊN, V. A. H.; CHAROENSOOK, S. N. Forecasting models for hydropower production using ARIMA method. In: **2021 9th international electrical engineering congress (IEECON)**. IEEE, p. 197-200, 2000

PONTES, L. P. **Previsão de séries temporais: produção industrial e demanda de energia elétrica residencial no Brasil**, 2018. Tese de Doutorado, DF: Universidade Católica de Brasília (UCB).

R CORE TEAM R: **A language and environment for statistical computing, R Foundation for Statistical Computing**, Viena, Áustria, 2020 Disponível em: Acesso em: abr. 2022.

RAPACH, D. E.; WEBER, C. E. Are real interest rates really nonstationary? New evidence from tests with good size and power. **Journal of Macroeconomics**, v. 26, n. 3, p. 409-430, 2004.

RINGKJØB, H. K.; HAUGAN, P. M.; SOLBREKKE, I. M. A review of modelling tools for energy and electricity systems with large shares of variable renewables. **Renewable and Sustainable Energy Reviews**, v. 96, p. 440-459, 2018.

SANTOS, J. B.; JABBOUR, C. J. C. Adoção da energia solar fotovoltaica em hospitais: revisando a literatura e algumas experiências internacionais. **Saúde e sociedade**, v. 22, p. 972-977, 2013.

SCHEIDT, F. *et al.* Data analytics in the electricity sector –A quantitative and qualitative literature review. **Energy and AI**, v. 1, p. 100009, 2020.

SCHWERT, G. W. Tests for unit roots: A Monte Carlo investigation. **Journal of Business & Economic Statistics**, v. 7, p. 147-159, 2002.

SILVEIRA, A. G. **Estudo da demanda de energia elétrica no Brasil**, 2017. Dissertação de Mestrado, RS: Universidade Federal do Rio Grande (FURG).

STERN, D. I.; BURKE, P. J.; BRUNS, S. B. **The impact of electricity on economic development: A macroeconomic perspective**. 2019.

SUMER, K. K.; GOKTAS, O.; HEPSAG, A. The application seasonal latent variable in forecasting electricity demand as an alternative method. **Energy policy**, v. 37, n. 4, p. 1317-1322, 2009

TAYLOR, J. W.; MENEZES, L. M.; MCSHARRY, P. E. A comparison of univariate methods for forecasting electricity demand up to a day ahead. **International journal of forecasting**, v. 22, n. 1, p. 1-16, 2006.

TEIXEIRA, C. A. *et al.* Modelagem bottom-up do consumo por uso final de eletricidade no setor residencial no Brasil. **Ambiente Construído**, v. 22, p. 113-131, 2022.

TRAPLETTI, A.; HORNIK, K.; LEBARON, B. tseries: Time series analysis and computational finance. R package version 0.10–49. 2021.

VELÁSQUEZ H. *et al.* Electricity demand forecasting using a SARIMA-multiplicative single neuron hybrid model. **Dyna**, v. 80, n. 180, p. 4-8, 2013.

WANG, Wen. **Stochasticity, nonlinearity and forecasting of streamflow processes**. Amsterdam: IOS Press, 2006. 210p.

WICKHAM, H.; CHANG, W.; WICKHAM, M. H. Package ‘ggplot2’. **Create elegant data visualisations using the grammar of graphics. Version**, v. 2, n. 1, p. 1-189, 2016.

YANG, Y. *et al.* Modelling a combined method based on ANFIS and neural network improved by DE algorithm: A case study for short-term electricity demand forecasting. **Applied Soft Computing**, v. 49, p. 663-675, 2016.

ZEILEIS, F. *et al.* strucchange: An R package for testing for structural change in linear regression models. **Journal of Statistical Software**, v. 7, n. 2, p. 1-38, 2002. Disponível em: <http://www.jstatsoft.org/v07/i02>