

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

O USO DOS MÉTODOS SARIMA E SARIMAX NA MODELAGEM DA DEMANDA MENSAL
DE ENERGIA ELÉTRICA NO ESTADO DO RIO GRANDE DO SUL

por

Tuane Proença Pereira

Dissertação para obtenção do Título de
Mestre em Modelagem Computacional

Rio Grande, abril, 2021.

Tuane Proença Pereira

“O USO DOS MÉTODOS SARIMA E SARIMAX NA MODELAGEM DA DEMANDA MENSAL DE ENERGIA ELÉTRICA NO ESTADO DO RIO GRANDE DO SUL”

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 Mestre. Área concentração: Modelagem Computacional.

Aprovado em 09/04/2021.

BANCA EXAMINADORA

Documento assinado digitalmente
 Viviane Leite Dias de Mattos
Data: 12/04/2021 11:32:20-0300
CPF: 301.650.080-91

Profª. Drª. Viviane de Leite Dias de Mattos

Orientadora - FURG



Documento assinado digitalmente
Vera do Carmo Comparsi de Vargas
Data: 12/04/2021 13:19:22-0300
CPF: 507.467.150-72
Verifique as assinaturas em <https://v.ufsc.br>

Profª. Drª. Vera do Carmo Comparsi de Vargas

Coorientadora - UFSC

Prof. Dr. André Andrade Longaray

Coorientador - FURG



Documento assinado digitalmente
Luiz Ricardo Nakamura
Data: 12/04/2021 16:37:15-0300
CPF: 366.099.768-48
Verifique as assinaturas em <https://v.ufsc.br>

Prof. Dr. Luiz Ricardo Nakamura

UFSC

Rio Grande - RS
2021

Dedico esse trabalho a minha amada vó Ondina, aos meus pais, Rita Berenici e Mauro Rogério, e a minha irmã, Kelly Cristiane, 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 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.

A minha co-orientadora, Vera do Carmo Comparsi de Vargas, por dedicar parte de seu tempo para contribuir em aprimorar este trabalho.

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

Aos meus colegas de mestrado, em especial aos que hoje posso chamar de amigos: Marta Leão e Thiago da Silveira.

A Franciele Nornberg e Luis Otavio S. Grillo, que apesar da distância, acompanharam esta árdua caminhada.

A Samanta Ramos e sua família por toda importância que representam em minha vida.

RESUMO

Inúmeros trabalhos mostram a importância do planejamento e previsão da demanda de energia elétrica cujo objetivo é manter um equilíbrio entre a produção e consumo deste produto. Para se aproximar dos padrões e comportamentos dessa variável e reduzir os erros de previsão, o que impacta diretamente em questões técnicas e financeiras, torna-se importante conhecer a sazonalidade nos estudos de previsão. Este trabalho tem como objetivo analisar o comportamento da sazonalidade da demanda mensal de energia elétrica no estado do Rio Grande do Sul para as três principais classes consumidoras: residencial, comercial e industrial. Dados históricos foram usados de janeiro de 2004 até dezembro de 2019. Inicialmente foi realizada uma análise exploratória utilizando técnicas gráficas e analíticas para resumir os dados, bem como o teste ANOVA com blocos e o teste de Friedman para avaliação da presença de sazonalidade, o que levou à escolha do modelo SARIMA, entre os disponibilizados pela metodologia de Box-Jenkins. A estacionariedade da parte sazonal e da parte não sazonal foram avaliadas pelo teste de Dickey-Fuller Aumentado. Após, os coeficientes de 256 modelos candidatos foram estimados pelo método de máxima verossimilhança, e o modelo mais parcimonioso determinado pelo critério de informação bayesiano. Na análise dos resíduos foram utilizados os testes de Ljung-Box, Jarque-Bera e ARCH, além dos indicadores MAE, RMSE, MAPE e estatística U de Theil para avaliar a capacidade preditiva, complementados por intervalos de confiança. O modelo encontrado em cada classe consumidora foi refinado com a inserção de variáveis exógenas: temperatura média, quantidade de consumidores e tarifa. Na modelagem SARIMAX, foram utilizados diagramas de dispersão e correlação cruzada para averiguar a relação entre estas variáveis e a demanda de energia elétrica, sendo utilizados os mesmos procedimentos para verificar a parcimonialidade, a adequação do ajuste e capacidade preditiva. Para a demanda mensal de energia elétrica na classe residencial o modelo selecionado é o SARIMA (1,1,2)(1,0,1) e na classe comercial é o modelo SARIMA (1,1,2)(0,1,1), ambos sem a inserção de variáveis exógenas. Na classe industrial o modelo selecionado é o SARIMA (1,1,1)(1,0,2)₁₂NC_{ind_3}, Tar_{ind_2}. Os modelos encontrados podem ser utilizados para prever a demanda mensal de energia elétrica no estado do Rio Grande do Sul, fornecendo subsídios para tomada de decisões deste setor. Também podem ser usados para elaboração de políticas públicas relacionadas ao desenvolvimento econômico da região. Além disso, no desenvolvimento do estudo, a aplicação de todos os métodos e técnicas utilizados foi acompanhada de uma análise crítica em relação ao seu desempenho, sendo detectadas algumas falhas. Estes achados podem servir para o desenvolvimento de futuros estudos que visem a melhorar o desempenho da metodologia utilizada, possibilitando modelos mais adequados e previsões mais precisas.

Palavras-chaves: Séries Temporais, Previsão, SARIMA, SARIMAX, Sazonalidade.

ABSTRACT

Numerous studies show the importance of planning and forecasting electricity demand, which aims to maintain a balance between production and consumption of this product. In order to approach the patterns and behaviors of this variable and reduce forecasting errors, that directly affects technical and financial issues, it is important to know the seasonality in forecasting studies. This work aims to analyze the behavior of the seasonality of the monthly demand for electricity in the Rio Grande do Sul state to the three main consumer classes: residential, commercial and industrial. Historical data were used from January 2004 until December 2019. Initially, an exploratory analysis was performed using graphical and analytical techniques to summarize the data, as well as the ANOVA test with blocks and the Friedman test to assess the presence of seasonality, which led to the choice of the SARIMA model, among those made available by the Box-Jenkins methodology. The stationarity of the seasonal part and the non-seasonal part were evaluated by the Augmented Dickey-Fuller test. Afterwards, the coefficients of 256 candidate models were estimated using the maximum likelihood method, and the most parsimonious model determined by the Bayesian information criterion. In the analysis of residues, the Ljung-Box, Jarque-Bera and ARCH tests were used, in addition to MAE, RMSE, MAPE and Theil's U statistics to assess predictive capacity, complemented by reliable intervals. The model found in each consumer class was refined with the inclusion of exogenous variables: average temperature, number of consumers and tariff. In the SARIMAX modeling, dispersion and cross-correlation diagrams were used to ascertain the relationship between these variables and the electricity demand, using the same procedures to check the parsimony, the adequacy of adjustment and the predictive capacity. For the monthly electricity demand in the residential class the model selected is SARIMA (1.1.2)(1.0.1) and in the commercial class it is the SARIMA model (1.1.2)(0.1, 1), both without the inclusion of exogenous variables. In the industrial class, the selected model is SARIMA (1,1,1)(1,0,2)12NCind_3,Tarind_2. The models found can be used to forecast the monthly electricity demand in the Rio Grande do Sul state, providing subsidies for decision making in this sector. They can also be used to develop public policies related to the economic development of the region. In addition, in the development of the study, the application of all methods and techniques used was accompanied by a critical analysis in relation to its performance, with some faults being detected. These findings may serve for the development of future studies that aim to improve the performance of the methodology used, allowing for more appropriate models and more accurate forecasts.

Keywords: Time Series, Forecast, SARIMA, SARIMAX, Seasonality.

ÍNDICE

1. INTRODUÇÃO	18
1.1. Apresentação do tema	18
1.2. Justificativa	21
1.3. Objetivo geral.....	22
1.3.1. Objetivos específicos.....	22
1.4. Estrutura do trabalho.....	23
2. REFERENCIAL TEÓRICO	24
2.1. Conceitos básicos de séries temporais	24
2.1.1. Estacionariedade.....	25
2.1.2. Ruído branco.....	25
2.1.3. Passeio aleatório	26
2.1.4. Operadores de retardo e diferença	26
Operador de retardo (B).....	27
Operador diferença (∇)	27
2.2. Modelos Box-Jenkins.....	28
2.2.1. Modelos ARMA (p,q).....	28
2.2.2. Modelos ARIMA (p,d,q).....	30
2.2.3. Modelos SARIMA (p,d,q)(P,D,Q) _s	30
2.2.4. Modelos SARIMAX (p,d,q)(P,D,Q) _s X	30
2.3. Etapas da modelagem SARIMA	31
2.3.1. Especificação	31
Análise exploratória dos dados	31
Verificação da sazonalidade.....	32
Verificação da estacionariedade.....	34
2.3.2. Identificação	36
2.3.3. Estimação dos parâmetros	38
Critério de Informação.....	39
Análise dos resíduos	40
2.3.4. Validação.....	42
3. METODOLOGIA	45
3.1. Procedimentos metodológicos da modelagem SARIMA	45
3.2. Procedimentos metodológicos para o refinamento do modelo: SARIMAX.....	47

4. RESULTADOS E DISCUSSÕES	49
4.1. Modelagem SARIMA	49
4.2. Refinamento do modelo: SARIMAX	73
4.3. Avaliação dos modelos desenvolvidos, métodos e técnicas utilizadas	97
5. CONSIDERAÇÕES FINAIS.....	99
REFERÊNCIAS.....	101
APÊNDICE 1 – Diagramas de dispersão entre as demandas mensais de energia elétrica e as variáveis exógenas.....	105
APÊNDICE 2 – Análise de estacionariedade das variáveis externas	108

LISTA DE FIGURAS

Figura 2.1 - Exemplos de gráficos da FAC e FACP para um modelo AR(1) e MA(1).....	38
Figura 3.1 - Esquema das etapas metodológicas para modelagem SARIMA	46
Figura 3.2 – Esquema das etapas metodológicas para modelagem SARIMAX.....	47
Figura 4.1 – Demanda mensal de energia elétrica da classe residencial/RS, 2004 - 19	49
Figura 4.2 - Demanda mensal de energia elétrica da classe comercial/RS, 2004 - 19.....	50
Figura 4.3 - Demanda mensal de energia elétrica da classe industrial/RS, 2004 - 19	50
Figura 4.4 - Representação do valor da tendência central para cada mês ao longo dos anos da demanda mensal de energia elétrica para a classe residencial/RS, 2014 - 19	51
Figura 4.5 - Representação do valor da tendência central para cada mês ao longo dos anos da demanda mensal de energia elétrica para a classe comercial/RS, 2004 - 19.....	51
Figura 4.6 - Representação do valor da tendência central para cada mês ao longo dos anos da demanda mensal de energia elétrica para a classe industrial/RS, 2004 - 19	52
Figura 4.7 - Gráfico de caixa para cada ano da demanda mensal de energia elétrica da classe residencial/RS, 2004 - 19.....	53
Figura 4.8 - Gráfico de caixa para cada ano da demanda mensal de energia elétrica da classe comercial/RS, 2004 - 19.....	54
Figura 4.9 - Gráfico de caixa para cada ano da demanda mensal de energia elétrica da classe industrial/RS, 2004 - 19	54
Figura 4.10 - Gráfico de caixa para cada mês ao longo dos anos da demanda mensal de energia elétrica da classe residencial/RS, 2004 – 19	55
Figura 4.11 - Gráfico de caixa para cada mês ao longo dos anos da demanda mensal de energia elétrica da classe comercial/RS, 2004 - 19	55
Figura 4.12- Gráfico de caixa para cada mês ao longo dos anos da demanda mensal de energia elétrica da classe industrial/RS, 2004 - 19.....	56
Figura 4.13 - Logaritmo da demanda mensal de energia elétrica da classe residencial/ RS, 2004 – 19	59
Figura 4.14 - Logaritmo da demanda mensal de energia elétrica da classe comercial/RS, 2004 – 19	59
Figura 4.15 - Logaritmo da demanda mensal de energia elétrica da classe industrial/RS, 2004 – 19	60
Figura 4.16 - FAC e FACP, da parte não sazonal e sazonal, do logaritmo da demanda mensal de energia elétrica da classe residencial/RS, 2004 - 18	61

Figura 4.17 - FAC e FACP, da parte não sazonal e sazonal, do logaritmo da demanda mensal de energia elétrica da classe comercial/RS, 2004 – 18	61
Figura 4.18 - FAC e FACP, da parte não sazonal e sazonal, do logaritmo da demanda mensal de energia elétrica da classe industrial/RS, 2004 - 18	62
Figura 4.19 - Gráfico do modelo de previsão SARIMA (1,1,2) (1,0,1) ₁₂ para demanda mensal de energia elétrica da classe residencial/RS, 2004 - 19	71
Figura 4.20 - Gráfico do modelo de previsão SARIMA (1,1,2) (0,1,1) ₁₂ para demanda mensal de energia elétrica da classe comercial/RS, 2004 - 19.....	71
Figura 4.21 - Gráfico do modelo de previsão SARIMA (1,1,1) (1,0,2) ₁₂ para demanda mensal de energia elétrica da classe industrial/RS, 2004 - 19	72
Figura 4.22 - Análise dos resíduos dos modelos selecionados para as três classes consumidoras/RS, 2004 - 18	73
Figura 4.23 - Número de consumidores da classe residencial (a) em linha (b) subséries sazonais/RS, 2004 - 19	74
Figura 4.24 - Número de consumidores da classe comercial (a) em linha (b) subséries sazonais/RS, 2004 - 19	74
Figura 4.25 - Número de consumidores da classe industrial (a) em linha (b) subséries sazonais/RS, 2004 - 19	75
Figura 4.26 - Tarifa mensal de energia elétrica para a classe residencial (a) em linha (b) subséries sazonais/RS, 2004 - 19.....	76
Figura 4.27 - Tarifa mensal de energia elétrica para a classe comercial (a) em linha (b) subséries sazonais/RS, 2004 - 19.....	76
Figura 4.28 - Tarifa mensal de energia elétrica para a classe industrial (a) em linha (b) subséries sazonais/RS, 2004 - 19.....	77
Figura 4.29 - Gráfico mensal da temperatura (a) em linha (b) subséries sazonais/RS, 2004 - 19	77
Figura 4.30 - Gráfico da correlação cruzada da parte não sazonal (a) e da parte sazonal (b) da demanda mensal de energia elétrica com o número de consumidores residenciais/RS, 2004 - 18.....	83
Figura 4.31 - Gráfico da correlação cruzada da parte não sazonal (a) e da parte sazonal (b) da demanda mensal de energia elétrica com a tarifa residencial/RS, 2004 - 18	83
Figura 4.32 - Gráfico da correlação cruzada da parte não sazonal (a) e da parte sazonal (b) da demanda mensal de energia elétrica da classe residencial com a temperatura/RS, 2004 - 18	84
Figura 4.33 - Gráfico da correlação cruzada da parte não sazonal (a) e da parte sazonal (b) da demanda mensal de energia elétrica com o número de consumidores comerciais/RS, 2004 – 18.....	85
Figura 4.34- Gráfico da correlação cruzada da parte não sazonal (a) e da parte sazonal (b) da demanda	

mensal de energia elétrica com a tarifa comercial/RS, 2004 – 18	85
Figura 4.35- Gráfico da correlação cruzada da parte não sazonal (a) e da parte sazonal (b) da demanda mensal de energia elétrica da classe comercial com a temperatura/RS, 2004 – 18	86
Figura 4.36 - Gráfico da correlação cruzada da parte não sazonal (a) e da parte sazonal (b) da demanda mensal de energia elétrica com o número de consumidores industriais/RS, 2004 – 18	87
Figura 4.37 - Gráfico da correlação cruzada da parte não sazonal (a) e da parte sazonal (b) da demanda mensal de energia elétrica com a tarifa industrial/RS, 2004 – 18.....	87
Figura 4.38 - Gráfico da correlação cruzada da parte não sazonal (a) e da parte sazonal (b) da demanda mensal de energia elétrica da classe industrial com a temperatura/RS, 2004 – 18.....	88
Figura 4.39 - Gráfico do modelo de previsão SARIMA (1,1,2)(1,0,1) ₁₂ Tar _{res} _1 para demanda mensal de energia elétrica da classe residencial/RS, 2004 - 19	96
Figura 4.40 - Gráfico do modelo de previsão SARIMA (1,1,1)(1,0,2) ₁₂ NC _{ind} _3,Tar _{ind} _2 para demanda mensal de energia elétrica da classe industrial/RS, 2004 - 19	96
Figura 0.1- Diagrama de dispersão da demanda mensal de energia elétrica da classe residencial com as variáveis exógenas/RS, 2004 - 18.....	105
Figura 0.2 - Diagrama de dispersão da demanda mensal de energia elétrica da classe comercial com as variáveis exógenas/RS, 2004 - 18.....	106
Figura 0.3 - Diagrama de dispersão da demanda mensal de energia elétrica da classe industrial com as variáveis exógenas/RS, 2004 - 18.....	107

LISTA DE TABELAS

Tabela 2.1- Tabela cruzada dos dados	32
Tabela 2.2 - Resumo do comportamento gráfico das funções FAC e FACP	37
Tabela 3.1 - Dados e fontes	45
Tabela 4.1 - Medidas resumo para as três classes consumidoras da demanda mensal de energia elétrica/RS, 2004 - 19.....	53
Tabela 4.2 - Resultado da ANOVA em blocos da demanda mensal de energia elétrica para as três classes consumidoras/RS, 2004 - 19	57
Tabela 4.3- Análise dos resíduos do teste ANOVA com blocos da demanda mensal de energia elétrica/RS, 2004 -19.....	58
Tabela 4.4 - Resultados do teste de Friedman da demanda mensal de energia elétrica para as três classes consumidoras/RS, 2004 - 19	58
Tabela 4.5 - Resultado do teste de estacionariedade da demanda mensal de energia elétrica para as três classes consumidoras/RS, 2004 - 18	60
Tabela 4.6 - Resultado do critério de informação e análise dos resíduos da demanda mensal de energia elétrica para classe residencial/RS, 2004 -18.....	63
Tabela 4.7- Resultado do critério de informação e análise dos resíduos da demanda mensal de energia elétrica para classe comercial/RS, 2004 - 18	64
Tabela 4.8 - Resultado do critério de informação e análise dos resíduos da demanda mensal de energia elétrica para classe industrial/RS, 2004 - 18	65
Tabela 4.9 - Indicadores de qualidade para validação dos modelos da classe residencial/RS, 2019	66
Tabela 4.10 - Indicadores de qualidade para validação dos modelos da classe comercial, 2019	66
Tabela 4.11 - Indicadores de qualidade para validação dos modelos da classe industrial/RS, 2019	67
Tabela 4.12 - Modelo SARIMA (1,1,2)(1,0,1) ₁₂ , valores observados, previstos e intervalo de confiança da demanda mensal de energia elétrica para a classe residencial/RS, 2019.....	68
Tabela 4.13 - Modelo SARIMA (1,1,2)(0,1,1) ₁₂ , valores observados, previstos e intervalo de confiança da demanda mensal de energia elétrica para a classe comercial/RS, 2019	69
Tabela 4.14 - Modelo SARIMA (1,1,1)(1,0,2) ₁₂ , valores observados, previstos e intervalo de confiança da demanda mensal de energia elétrica para a classe industrial/RS, 2019	70
Tabela 4.15 - Medidas de resumo dos dados mensais das variáveis exógenas para as três classes consumidoras/RS, 2004 - 19	79
Tabela 4.16 - Resultado da ANOVA com blocos das variáveis externas para as três classes consumidoras/RS, 2004 - 19	80

Tabela 4.17 - Análise dos resíduos do teste ANOVA com blocos das variáveis externas para as três classes consumidoras/RS, 2004 - 19	81
Tabela 4.18 - Resultado do teste de Friedman das variáveis externas das três classes consumidoras/RS, 2004 - 19	82
Tabela 4.19 – Resultado do critério de informação do modelo residencial SARIMA (1,1,2)(1,0,1) ₁₂ com cada umas das variáveis exógenas/RS, 2004 - 18	89
Tabela 4.20 – Resultado do critério de informação do modelo residencial SARIMA (1,1,2)(1,0,1) ₁₂ com combinações das variáveis exógenas/RS, 2004 - 18	89
Tabela 4.21- Resultado do critério de informação do modelo comercial SARIMA (1,1,2)(0,1,1) ₁₂ com cada umas das variáveis exógenas/RS, 2004 - 18	90
Tabela 4.22 - Resultado do critério de informação do modelo industrial SARIMA (1,1,1)(1,0,2) ₁₂ com cada umas das variáveis exógenas/RS, 2004 - 18	91
Tabela 4.23 - Resultado do critério de informação do modelo industrial SARIMA (1,1,1)(1,0,2) ₁₂ e a combinação das variáveis exógenas/RS, 2004 - 18	91
Tabela 4.24 - Resultado da análise dos resíduos dos modelos SARIMAX nas três classes consumidoras/RS, 2004 - 18	92
Tabela 4.25 - Indicadores de qualidade para validação dos modelos SARIMAX das três classes/RS, 2019.....	93
Tabela 4.26 - Valores observados, previstos e intervalo de confiança do modelo residencial SARIMA (1,1,2)(1,0,1) ₁₂ Tar _{res} _1/RS, 2019	94
Tabela 4.27 - Valores observados, previstos e intervalo de confiança do modelo industrial SARIMA (1,1,1)(1,0,2) ₁₂ NC _{ind} _3,Tar _{ind} _2/ RS, 2019	95
Tabela 0.1 - Resultado do teste de estacionariedade para as variáveis exógenas/RS, 2004 - 18	108

LISTA DE SÍMBOLOS

A	Assimetria
B	Operador de retardo
d	Número de diferenças
D	Número de diferenças sazonais
$E[.]$	Esperança matemática
g	Meses de uma série temporal
H	Número total de observações no período de validação do modelo
h	Anos de uma série temporal
K	Curtose
M_d	Mediana
p	Número de defasagens de um processo autorregressivo
P	Número de defasagens de um processo autorregressivo sazonal
q	Número de defasagens de um processo de médias móveis
Q	Número de defasagens de um processo de médias móveis sazonal
QM	Quadrado médio
R	<i>Ranks</i>
SQ	Somas de quadrados
T	Número total de observações da série temporal / tamanho da amostra
U	Estatística U de Theil
var	Variância
W	Vetor de parâmetros populacionais da função de máxima verossimilhança
$X_{k,t}$	Observações de variáveis externas relacionadas às variáveis dependentes
Y_t	Série temporal
\tilde{Y}_t	Valor previsto

Símbolos gregos

σ^2	Variância
μ	Média
β_1	Coefficiente de constante no teste ADF
β_2	Coefficiente de tendência no teste ADF
γ	Autocovariância
δ	Coefficiente de presença de raiz unitária no teste ADF
ε_t	Ruído branco
Θ_Q	Coefficiente de médias móveis sazonal
θ_q	Coefficiente de médias móveis
ρ_j	Função de autocorrelação
ϕ_p	Coefficiente autorregressivo
ψ_k	Coefficiente de regressão das variáveis exógenas
Φ_P	Coefficiente autorregressivo sazonal
∇	Operador diferença

Sub índices

t	Instante de tempo
k	Defasagem temporal

LISTA DE ABREVIATURAS

ACL	Ambiente de Contratação Livre
ACR	Ambiente de Contratação Regulada
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
AR	Autorregressivo
ARIMA	Autoregressivo integrado e de média móvel
ARMA	Autorregressivo com média móvel
BIC	Critério de informação Bayesiano
Dem _{com}	Demanda mensal de energia elétrica da classe comercial
Dem _{ind}	Demanda mensal de energia elétrica da classe industrial
Dem _{res}	Demanda mensal de energia elétrica da classe residencial
EPE	Empresa de Pesquisa Energética
FAC	Função de autocorrelação
FACP	Função de autocorrelação parcial
IC	Intervalo de confiança
INMET	Instituto Nacional de Meteorologia
IPCA	Índice Nacional de Preços ao Consumidor Amplo
JB	Jarque-Bera
LB	Ljung-Box
MA	Médias móveis
MAE	Erro Absoluto Médio
MAPE	Erro Médio Percentual Absoluto

NC _{com}	Número de consumidores da classe comercial
NC _{ind}	Número de consumidores da classe industrial
NC _{res}	Número de consumidores da classe residencial
PIB	Produto Interno Bruto
RB	Ruído branco
RMSE	Raiz do Erro Quadrático Médio
SARIMA	Autoregressivo integrado e de média móvel sazonal
SARIMAX	Autorregressivo integrado e de média móvel sazonal com fator exógeno
SIN	Sistema Interligado Nacional
Tar _{com}	Tarifa mensal de energia elétrica para classe comercial
Tar _{ind}	Tarifa mensal de energia elétrica para classe industrial
Tar _{res}	Tarifa mensal de energia elétrica para classe residencial
Temp	Temperatura
VAR	Vetores autoregressivos
VAR-VEC	Vetores autoregressivo em conjunto com Correções de Erros Vetoriais

1. INTRODUÇÃO

1.1. Apresentação do tema

A energia elétrica é um item considerado imprescindível para a sobrevivência humana e para a qualidade de vida do homem moderno, fazendo-se presente nas mais variadas áreas da sociedade. Tal é a sua importância que se torna difícil de imaginar a humanidade sem a presença de energia elétrica (NUNES, 2019).

Uma particularidade do setor elétrico é que a estocagem de energia se torna inviável não só por razões técnicas, mas financeiras. Dessa maneira, é fundamental existir um planejamento do setor energético para que a demanda de energia elétrica seja a mais estável possível, em outras palavras, deve-se minimizar ao máximo os riscos de desequilíbrios entre produção e consumo de energia elétrica (THOMAZ, 2017). Quando a produção é maior que o consumo se tem um desperdício de energia, e quando o consumo é maior que a produção, é provável que ocorram problemas como a chamada crise do apagão, ocorrida em 2001 e 2002 no Brasil (SILVEIRA, 2017).

De acordo com Lima e Bacchi (2019), o planejamento e o gerenciamento da demanda de energia elétrica no Brasil são estruturais e circunstancialmente importantes, pois a futura demanda de energia elétrica afeta diretamente ou indiretamente as decisões das empresas envolvidas em sua produção, transmissão e distribuição. Para elaborar programas de expansão de geração de energia elétrica a longo prazo, é necessário um planejamento dos investimentos, projetos de infraestruturas para redes de transmissão e distribuição, além de avaliação do preço da energia no mercado, entre outros. Em outras palavras, realizar análises precisas da futura demanda de energia elétrica reduz incertezas relacionadas ao desenvolvimento do mercado consumidor.

Segundo Cabral et al. (2017), o Sistema Elétrico Brasileiro, com o objetivo de garantir eficiência energética, autonomia econômica e atender a demanda total de energia, passou por duas grandes reformas que culminaram no atual modelo elétrico institucional. Dentro dessas mudanças, a comercialização de energia no Brasil passa a ser realizada por duas esferas de mercado: Ambiente de Contratação Regulada (ACR) e o Ambiente de Contratação Livre (ACL). No grupo da ACL, estão presentes os consumidores “livres”, que atendendo alguns requisitos técnicos, podem comprar eletricidade de qualquer fornecedor, o que difere dos consumidores presentes no ACR, chamados de consumidores “cativos”, que pagam pelo valor da energia através de contratos firmados em leilões.

No que se refere à comercialização da energia elétrica, as empresas do serviço público, que assinam os contratos de fornecimento, devem estar seguras da sua capacidade em atender as demandas do mercado ou estarão sujeitas a penalidades. Cabral et al. (2017) salientaram que se as

previsões de demanda estiverem erradas, acima ou abaixo de 3% da demanda real, as empresas pagam multas muitas elevadas. Dessa maneira, os autores destacaram que é fundamental conhecer o comportamento de consumo e utilizar boas ferramentas para realizar as previsões da demanda de energia.

Para Elamin e Fukushige (2018), os custos financeiros dos erros de previsão são tão altos que muitas pesquisas estão focadas em reduzir o erro, mesmo que essa redução seja em frações percentuais. Estes autores, realizaram a projeção da demanda de energia elétrica para uma região do Japão, com os dados medidos de hora em hora entre 1º de janeiro de 2012 e 31 de dezembro de 2015. Desses quatro anos, três foram empregados para fase de teste e um ano para validação do modelo. Os modelos utilizados foram Autorregressivo Integrado de Média Móvel Sazonal com fator exógeno (modelo SARIMAX), analisando os efeitos principais das variáveis climáticas (temperatura e umidade) e variáveis simuladas para sazonalidade diária (horários de pico e fora de pico), semanal (dias úteis da semana) e anual (meses), e respectivas interações. A qualidade dos modelos foi avaliada utilizando o erro percentual absoluto médio (MAPE), o erro médio absoluto (MAE) e o erro médio quadrático da raiz (RMSE), sendo que o modelo SARIMAX com interações apresentou melhor desempenho quando comparado ao modelo de Regressão Linear Múltipla (RLM) com interações e ao modelo SARIMA.

Em função do exposto, a modelagem e previsão da demanda de energia elétrica se tornou um instrumento importante para o planejamento do setor elétrico, em que são aplicadas técnicas de análise de séries temporais, identificando estruturas e padrões de comportamento de sua demanda ao longo dos anos, e assim construir modelos que tenham capacidade de gerar boas previsões e auxiliar nas tomadas de decisões (PONTES, 2018).

Além disso, outro ponto importante a ser destacado sobre a demanda de energia, é que a mesma serve como um índice de desenvolvimento econômico de um país (NUNES, 2019). Segundo Silveira (2017), países mais desenvolvidos apresentam maior consumo de energia elétrica, se comparados a países em desenvolvimento. Dessa forma, a modelagem da demanda de energia, além de contribuir para um bom planejamento energético, também pode auxiliar nas projeções econômicas de um país.

Hyndman e Athanasopoulos (2018) apresentam alguns métodos que podem ser utilizados na modelagem de séries temporais, como: Regressão, Decomposição de séries temporais, Suavização Exponencial e ARIMA, entre outros. Dentre esses métodos, a aplicação da metodologia Box Jenkins no setor energético para modelagem da demanda de energia elétrica é bastante consolidada como mostram os seguintes trabalhos.

Pereira (2015) compararam o desempenho de dois métodos (SARIMAX e Sistema de Inteligência Fuzzy - ANFIS) para prever a distribuição de energia elétrica em uma região localizada na Bahia, utilizando diferentes combinações de três variáveis exógenas: temperatura ambiente, precipitação pluviométrica e número de clientes. Os modelos obtidos por ANFIS apresentaram melhor desempenho, que foi diretamente proporcional ao aumento da quantidade de variáveis inseridas. Aumentar a quantidade de variáveis exógenas não forneceu aos modelos SARIMAX uma melhor capacidade preditiva.

Tarsitano e Amerise (2017) desenvolverem estudo para prever a demanda horária de energia elétrica em seis regiões da Itália. Para modelar as 27.972 informações obtidas entre maio de 2013 e junho de 2016, foi utilizada uma modelagem híbrida composta por regressão, seguida por modelagem SARIMA aplicada aos seus resíduos e, se necessário, SARIMAX. Os autores consideraram que a principal contribuição do estudo foi ter desenvolvido um modelo novo de regressão dinâmica para explicar as mudanças na demanda de energia elétrica em função de mudanças em variáveis explicativas. Ainda, consideraram que o método proposto foi eficiente para previsões de curto prazo, além de não ser muito exigente em termos de recursos computacionais.

Além dos setores de energia, a metodologia de Box-Jenkins, mais especificamente as modelagens SARIMA e SARIMAX, tem sido também bastante utilizada com sucesso em diversas áreas da ciência. Espinosa et al. (2010), por exemplo, a utilizaram para descrever o comportamento e fazer previsões do número de focos de calor no estado de Mato Grosso. Foram utilizados dados mensais de junho a outubro por um período de 17 anos, sendo identificado o modelo SARIMA $(1,0,0)(0,1,1)_5$ como aquele de melhor desempenho. Os autores salientaram a importância do modelo encontrado por permitir a determinação de medidas preventivas para a diminuição de focos de calor e, dessa forma, ajudar a diminuir os danos causados, principalmente, ao meio ambiente.

Já Pinto et al. (2018) desenvolveram um estudo com o objetivo de modelar e prever a concentração média diária de material particulado inalável de uma região no estado do Espírito Santo. Além de dados mensais da variável a ser prevista por um período de aproximadamente três anos, utilizaram dados de variáveis ambientais que poderiam ou não entrar no modelo como variáveis exógenas. Como a série da variável dependente foi considerada estacionária em nível na parte não sazonal, foram utilizados os modelos ARMA, SARMA, ARMAX, SARMAX e SARIMAX, identificando o modelo SARIMAX $(1,0,2)(0,1,1)_7$ como aquele que apresenta melhor capacidade preditiva de acordo com o RMSE e MAE. Os autores deixam como sugestão, a utilização de um modelo híbrido composto pelo modelo de heterocedasticidade condicional autorregressivo generalizado (GARCH) e SARIMAX.

1.2. Justificativa

Diversos estudos com dados nacionais, que realizaram a modelagem da demanda de energia elétrica, podem ser encontrados na literatura. Lima e Bacchi (2019) analisaram os efeitos do aumento do Produto Interno Bruto (PIB) na demanda de energia elétrica no Brasil, entre os anos de 2004 e 2018. Para isso realizaram uma modelagem tendo como base dados de uma série temporal com dados mensais do PIB brasileiro e a demanda de energia elétrica total do Brasil, sem segmentação regional (dados nacionais) e sem classificação do tipo de consumidor (consumo total da energia no Brasil). De acordo com as autoras, o aumento de 1% no PIB, considerado proxy do crescimento econômico, acarreta no aumento de 0,475% na demanda total de energia elétrica do Brasil.

Silveira (2017) desenvolveu modelos para previsões da demanda de energia elétrica do setor brasileiro, cuja base de dados é a demanda anual de energia elétrica do Brasil. Já Nunes (2019), empregando dados anuais da demanda no estado do Rio Grande do Sul, elaborou modelos para previsões da demanda deste estado. Em ambos os estudos foram feitas as modelagens distintas para cada uma das três principais classes de consumidores do sistema elétrico: residencial, comercial e industrial. De acordo Silveira (2017), estudos mostram que cada classe de consumidores utiliza a energia elétrica de forma distinta, logo é usual que as modelagens ocorram separadamente para cada setor.

Oliveira et al. (2001) estudaram a diversidade sazonal da demanda de energia elétrica no Brasil, com objetivo de mostrar as características dinâmicas da demanda e assim verificar as possibilidades para substanciais ganhos de eficiência econômica. Isto possibilitou utilizar seus resultados em estudos econômicos que analisem o trade-off entre maior capacidade de transporte e maior capacidade de oferta nos centros de carga. Estes autores utilizaram dados mensais, do período de 1976 a 1997, divididos pelas cinco regiões: Norte, Sul, Sudeste, Nordeste e Centro-Oeste, e pelas diversas classes de consumo: industrial, residencial, comercial, rural, poderes públicos, iluminação pública, serviços públicos e demais classes. A técnica empregada foi a decomposição sazonal denominada Seasonal-Trend Decomposition Procedure Based on Loess (STL), a fim de encontrar as componentes de tendência e sazonalidade da série investigada. As principais conclusões do seu estudo foram a existência de uma componente sazonal significativa na demanda de energia elétrica considerando uma análise nacional, embora esta não tenha sido estática. Com relação às observações das classes consumidoras e às cinco regiões analisadas, verificaram-se sazonalidades bastante diferentes entre si. Além disso, concluíram que o sistema elétrico é capaz de atender a demanda de energia, mesmo quando ocorrem picos, o que é desejável.

Realizar a modelagem da demanda de energia acrescido do conhecimento sobre a

sazonalidade desta série é bastante relevante nos estudos de previsão da demanda de energia elétrica, pois permite encontrar modelos com ajuste bastante satisfatório, reduzindo os erros de previsão, o que é muito importante tecnicamente e financeiramente. Com relação ao Brasil, estudos como o desenvolvido por Oliveira et al. (2001), permitem a formulação de políticas energéticas para melhorar o aproveitamento de energia.

Algumas dessas políticas já foram implementadas pela Agência Nacional de Energia Elétrica (ANEEL). Uma das mais conhecidas é o sistema tarifário de bandeira, lançada em 1º de maio de 2015, criada com o objetivo de reduzir a demanda de energia elétrica, atingindo todos os consumidores do Sistema Interligado Nacional (SIN) (ANEEL, 2015). E em 1º de janeiro de 2020 a ANEEL lançou a Tarifa Branca para os consumidores residenciais com o objetivo de aumentar a demanda de energia elétrica fora do horário de ponta, sendo que essa e outras modalidades tarifárias já são oferecidas a outros tipos de consumidores, devido ao seu perfil de consumo. (ANEEL, 2020)

Embora inúmeros estudos referentes a modelagem da demanda de energia elétrica no Brasil tenham sido desenvolvidos, em que foi possível a formulação de políticas públicas, até o presente momento estudos específicos sobre a sazonalidade dos dados da demanda de energia elétrica para o estado do Rio Grande do Sul e a obtenção de modelos de previsão com bases nesses dados não foram encontrados na literatura acadêmica, justificando a importância da realização deste estudo.

1.3. Objetivo geral

O presente estudo tem como objetivo analisar o comportamento da sazonalidade na demanda mensal de energia elétrica do estado do Rio Grande do Sul para as três principais classes consumidoras: residencial, comercial e industrial.

1.3.1. Objetivos específicos

- Realizar um estudo exploratório dos dados de demanda mensal de energia elétrica nas três principais classes consumidoras para verificar a existência da sazonalidade nos dados analisados;
- modelar a demanda mensal de energia elétrica do estado do Rio Grande do Sul para as três principais classes consumidoras: residencial, comercial e industrial;
- analisar o desempenho de algumas técnicas estatísticas utilizadas pela metodologia de Box-Jenkins na modelagem de séries temporais com sazonalidade;
- verificar a possibilidade de incluir variáveis explicativas nos modelos encontrados para melhorar o desempenho do modelo desenvolvido.

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 desenvolvimento deste trabalho. A metodologia do trabalho encontra-se no capítulo três. Nele mostra-se como foi realizada a manipulação dos dados para cada um dos modelos aplicados. No capítulo quatro são apresentados os resultados e discussão dos modelos encontrados em cada uma das três classes consumidoras analisadas. E por fim, no capítulo cinco encontram-se as considerações finais sobre o estudo realizado.

2. REFERENCIAL TEÓRICO

Neste capítulo são apresentados os principais conceitos básicos sobre séries temporais. Em seguida, são expostos alguns conceitos relacionados à modelagem de séries temporais, seguidos de considerações teóricas sobre as diversas etapas das construções dos modelos de Box & Jenkins: identificação, estimação, diagnóstico e validação. Por fim, são apresentados estudos similares.

2.1. Conceitos básicos de séries temporais

Um processo estocástico pode ser definido a partir de uma família $Y = \{Y_t, t \in T\}$, tal que, para cada $t \in T$, existe um $Y(t)$ que é uma variável aleatória, ou seja, processo estocástico é o conjunto de todos os possíveis valores de uma variável que pode ser observado no instante de tempo analisado (MORETTIN E TOLOI, 2018; BUENO, 2011).

A definição de série temporal pode ser compreendida como um conjunto de amostras referentes às realizações de um processo estocástico, em outras palavras, série temporal é uma sequência de valores observados de uma determinada variável em intervalos de tempo, muitas vezes, igualmente espaçados (por hora, por dia, por mês, por ano). A demanda de energia elétrica mensal, o resultado de vendas trimestrais de uma determinada empresa, bem como seus lucros anuais, são alguns exemplos de séries temporais. (GUJARATI e PORTER, 2011; HYNDMAN e ATHANASOPOULOS, 2018; BUENO, 2011).

Outra abordagem importante dentro do estudo de séries temporais é com relação aos padrões que as mesmas podem apresentar. Hyndman e Athanasopoulos (2018) consideram tendência (T_t), sazonalidade (S_t) e ciclo (C_t) como os três padrões que podem estar presentes em uma série. A tendência está relacionada com o aumento ou diminuição em longo prazo dos dados observados, sem a necessidade de ser linear. A sazonalidade está presente em séries com observações horárias, semanais, mensais ou trimestrais, sendo afetada por fatores sazonais, como: época do ano (efeitos climáticos), dias da semana (efeitos de calendário) ou horas do dia. Já o padrão de ciclo difere da sazonalidade, pois as oscilações existem sem uma frequência fixa e, em geral, estão ligadas a questões econômicas. Alguns autores ainda identificam um outro padrão, os movimentos irregulares (M_t), que são movimentos provindos de eventos aleatórios imprevisíveis, como por exemplo, uma pandemia.

Muitas séries apresentam os padrões descritos acima, logo, antes de escolher um modelo de previsão, é importante identificá-los e, posteriormente, escolher um método capaz de manipular esses componentes e realizar previsões adequadas. (MORETTIN e TOLOI, 2018; HYNDMAN e ATHANASOPOULOS, 2018).

2.1.1. Estacionariedade

A verificação da estacionariedade de uma série temporal é uma importante propriedade a ser considerada na modelagem de um processo estocástico. As séries temporais podem ser estritamente estacionárias ou fracamente estacionárias. Em geral, quando se verifica a estacionariedade de uma série, faz-se referência a estacionariedade fraca, também conhecida como estacionariedade de segunda ordem. (MORETTIN e TOLOI, 2018; BUENO, 2011; GUJARATI e PORTER, 2011).

Uma série é dita fracamente estacionária quando a média é igual para todo período, mesmo que a condição da distribuição da variável altere-se ao longo do tempo (Eq. 2.1.1.1); a variância seja sempre igual para todo período (Eq. 2.1.1.2); e a covariância não dependa do tempo, mas apenas da defasagem temporal (distância temporal) entre as observações (Eq. 2.1.1.3). Então, se uma série temporal for estacionária, a média, a variância e as autocovariâncias (para defasagens específicas) permanecerão as mesmas não importando em que ponto seja analisado, sendo então invariantes no tempo. (MORETTIN e TOLOI, 2018; BUENO, 2011; GUJARATI e PORTER, 2011).

$$E(Y_t) = \mu, \forall t \in T \quad (2.1.1.1)$$

$$var(Y_t) = E(Y_t - \mu)^2 = \sigma^2, \forall t \in T \quad (2.1.1.2)$$

$$\gamma_k = E[(Y_t - \mu)(Y_{t+k} - \mu)], \text{ para } k = \pm 1, \pm 2, \dots \quad (2.1.1.3)$$

Para fins de previsão, séries não estacionárias são de pouco valor prático, logo, caso a série original não seja estacionária, é possível realizar uma diferenciação até que a mesma se torne estacionária. Normalmente, depois da primeira ou segunda diferença a série se torna estacionária. (MORETTIN e TOLOI, 2018; BUENO, 2011; GUJARATI e PORTER, 2011).

2.1.2. Ruído branco

Considera-se ruído branco, representado por $\varepsilon_t \sim \text{RB}(0, \sigma^2)$, uma sequência de valores independente e identicamente distribuídos (i.i.d) de uma variável aleatória com média zero (Eq. 2.1.2.1), variância constante (Eq. 2.1.2.2) e que não esteja correlacionada a qualquer realização da própria série, ou seja, autocorrelação igual a zero (Eq. 2.1.2.3). (BUENO, 2011; GUJARATI e PORTER, 2011).

$$E(\varepsilon_t) = \mu_\varepsilon = 0, \forall t \in T \quad (2.1.2.1)$$

$$var(\varepsilon_t) = \sigma_\varepsilon^2, \forall t \in T \quad (2.1.2.2)$$

$$cov(\varepsilon_t, \varepsilon_{t+k}) = 0 \text{ sendo } k \neq 0 \quad (2.1.2.3)$$

Segundo Morettin e Tolo (2018), outra representação dada ao ruído branco, também chamado de processo puramente aleatório, é $\varepsilon_t \sim \text{RB}(0, \sigma_\varepsilon^2)$. Além disso, pelas características apresentadas, o ruído branco é um processo estacionário.

2.1.3. Passeio aleatório

Passeio aleatório pode ser definido pela Eq. 2.1.3.1, em que Y_t é igual a seu valor no tempo ($t-1$) mais um choque aleatório ε_t . (GUJARATI e PORTER, 2011).

$$Y_t = Y_{t-1} + \varepsilon_t \quad (2.1.3.1)$$

Reescrevendo a Eq. 2.1.3.1, para $t=1,2,3$, tem-se:

$$\begin{aligned} Y_1 &= Y_0 + \varepsilon_1 \\ Y_2 &= Y_1 + \varepsilon_2 = Y_0 + \varepsilon_1 + \varepsilon_2 \\ Y_3 &= Y_2 + \varepsilon_3 = Y_0 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3 \end{aligned}$$

Partindo da premissa que o processo se iniciou em algum tempo 0 com valor de Y_0 , recursivamente tem-se o passeio aleatório como apresentado na Eq. 2.1.3.2.

$$Y_t = Y_0 + \sum_{j=1}^t \varepsilon_j \quad (2.1.3.2)$$

Segundo Gujarati e Porter (2011), o passeio aleatório apresenta a persistência de choques aleatórios, conforme mostra a Eq. 2.1.3.2. Nota-se que Y_t é a soma de Y_0 com a soma dos choques aleatórios. Devido a essa característica, diz-se que o passeio aleatório tem memória infinita.

2.1.4. Operadores de retardo e diferença

A seguir são apresentados os principais operadores utilizados em estudos de séries temporais. O primeiro facilita a operacionalização matemática na modelagem dos dados, enquanto o segundo, a

obtenção da estacionariedade.

Operador de retardo (B)

O operador de retardo (B), também chamado de *backshift* ou de translação para o passado, tem como função retardar uma certa quantidade de períodos da série temporal, ou seja, representa a defasagem de k períodos passados. Este operador é mostrado na Eq. 2.1.4.1 para a primeira defasagem (MORETTIN e TOLOI, 2018; CRYER e CHAN, 2008).

$$BY_t = Y_{t-1} \quad (2.1.4.1)$$

A segunda defasagem da série temporal é apresentada na Eq. 2.1.4.2.

$$B^2Y_t = B(BY_t) = Y_{t-2} \quad (2.1.4.2)$$

Para k defasagens tem-se a Eq. 2.1.4.3.

$$B^kY_t = Y_{t-k} \quad (2.1.4.3)$$

Ao aplicar este operador com potência k na série temporal no tempo t (Y_t), obtém-se o valor desta série no tempo $t - k$.

Operador diferença (∇)

Este operador, representado por ∇ , realiza a diferença entre observações consecutivas de uma série temporal com t observações, resultando numa nova série com $t-1$ observações (CRYER e CHAN, 2008; BUENO, 2011). A Eq. 2.1.4.4 mostra a aplicação da primeira diferença nos dados de uma série temporal.

$$\nabla Y_t = (1 - B)Y_t = Y_t - Y_{t-1} \quad (2.1.4.4)$$

Aplicando-se o operador novamente, obtém-se a segunda diferença, representada na Eq. 2.1.4.5.

$$\nabla^2 Y_t = \nabla(\nabla Y_t) = \nabla(Y_t - Y_{t-1})$$

$$\nabla^2 Y_t = (1 - B)^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2} \quad (2.1.4.5)$$

A n -ésima diferença de Y_t é obtida recursivamente pela Eq. 2.1.4.6.

$$\nabla^n Y_t = (1 - B)^n Y_t \quad (2.1.4.6)$$

2.2. Modelos Box-Jenkins

Segundo Gujarati e Porter (2011), os modelos de Box-Jenkins são modelos desenvolvidos a partir de dados passados da própria série temporal, que somados a termos de erros estatísticos podem realizar previsões. Nesta seção são apresentados os modelos de Box-Jenkins, detalhando algumas especificidades.

2.2.1. Modelos ARMA (p,q)

O modelo autorregressivo com médias móveis (ARMA) é formado a partir da combinação de outros dois modelos: o modelo autorregressivo (AR) e o modelo de médias móveis (MA). O modelo autorregressivo, apresentado na Eq. 2.2.1.1 e representado por AR(p), é empregado quando for possível assumir que a determinação do valor atual da série depende dos valores passados da mesma (GUJARATI e PORTER, 2011), ou seja:

$$\tilde{Y}_t = \phi_1 \tilde{Y}_{t-1} + \phi_2 \tilde{Y}_{t-2} + \dots + \phi_p \tilde{Y}_{t-p} + \varepsilon_t \quad (2.2.1.1)$$

em que: $\tilde{Y}_t = Y_t - \mu$, sendo Y_t o valor observado, no instante de tempo t , μ é a média do processo, ϕ_i são os coeficientes autorregressivos que descrevem como o valor atual se relaciona com os valores anteriores ($i = 1, 2, \dots, p$), p representa a defasagem máxima de Y_t , ou seja, a quantidade máxima de períodos anteriores que é incluída no modelo, e $\varepsilon_t \sim R.B(0, \sigma_\varepsilon^2)$ um erro aleatório independente, com média zero e variância constante (ruído branco). (MORETTIN e TOLOI, 2018; BUENO, 2011; GUJARATI e PORTER, 2011).

Segundo Morettin e Toloi (2018), este modelo também pode ser representado através do emprego do operador de retardo, conforme mostra Eq. 2.2.1.2.

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \tilde{Y}_t = \varepsilon_t \quad (2.2.1.2)$$

Semelhante ao modelo $AR(p)$, o modelo de médias móveis, apresentado na Eq. 2.2.1.3 e representado por $MA(q)$, caracteriza-se por empregar os resíduos dos valores passados e atual da série temporal para predizer seu valor corrente (MORETTIN e TOLOI, 2018; BUENO, 2011; GUJARATI e PORTER, 2011), ou seja:

$$\tilde{Y}_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q} \quad (2.2.1.3)$$

em que: θ_j é o coeficiente de médias móveis que relaciona os erros passados ao valor atual observado ($j = 1, 2, \dots, q$), e q representa a defasagem máxima de ε_t , ou seja, quantidade máxima de erros dos períodos anteriores inseridos na série. Morettin e Toloi (2018) apresentam o modelo $MA(q)$ através do operador de retardo conforme Eq. 2.2.1.4.

$$\tilde{Y}_t = (1 - \theta_1B - \theta_2B^2 - \dots - \theta_qB^q)\varepsilon_t \quad (2.2.1.4)$$

Embora os dois modelos anteriores sejam bastante úteis, é provável que uma série temporal possa apresentar características que envolvam termos autorregressivos $AR(p)$ e termos de médias móveis $MA(q)$ simultaneamente. Logo, para atender séries desse tipo, utiliza-se o modelo autorregressivo com médias móveis, denominado como $ARMA(p,q)$, mostrado na Eq. 2.2.1.5 (MORETTIN e TOLOI, 2018; GUJARATI e PORTER, 2011).

$$\tilde{Y}_t = \phi_1\tilde{Y}_{t-1} + \dots + \phi_p\tilde{Y}_{t-p} - \theta_1\varepsilon_{t-1} - \dots - \theta_q\varepsilon_{t-q} + \varepsilon_t \quad (2.2.1.5)$$

O modelo $ARMA(p,q)$ também pode ser representado com o emprego dos operadores de retardo, como mostra a Eq. 2.2.1.6. (MORETTIN e TOLOI, 2018).

$$(1 - \phi_1B - \phi_2B^2 - \dots - \phi_pB^p)\tilde{Y}_t = (1 - \theta_1B - \theta_2B^2 - \dots - \theta_qB^q)\varepsilon_t \quad (2.2.1.6)$$

De maneira resumida, o modelo $ARMA(p,q)$ é a combinação do modelo $AR(p)$, relação entre os valores passados com o valor atual, e do modelo $MA(q)$, relação dos erros passados com o valor atual, acrescentada de um erro aleatório de média zero e variância constante. Na prática, muitas vezes, o uso do modelo $ARMA(p,q)$ se torna uma solução mais adequada para as séries temporais do que empregar apenas o modelo $AR(p)$ ou modelo $MA(q)$, tendo como restrição para uso desses modelos

a série ser estacionária. (MORETTIN e TOLOI, 2018; GUJARATI e PORTER, 2011).

2.2.2. Modelos ARIMA (p,d,q)

Quando as séries temporais são não estacionárias, utiliza-se o modelo autorregressivo integrado de médias móveis, denominado ARIMA (p,d,q) como mostra a Eq. 2.2.2.1 (MORETTIN e TOLOI, 2018; GUJARATI e PORTER, 2011).

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d Y_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t \quad (2.2.2.1)$$

O modelo ARIMA (p,d,q) é semelhante ao modelo anterior, porém acrescentado do termo $(1-B)^d$ referente à integração da série temporal para obter a estacionariedade, em que d representa a ordem de integração, ou seja, número de diferenças necessárias para que a série não estacionária satisfaça a condição de estacionariedade. (MORETTIN e TOLOI, 2018).

2.2.3. Modelos SARIMA (p,d,q)(P,D,Q)_s

Para séries temporais com características sazonais utiliza-se o modelo ARIMA sazonal, também chamado de SARIMA(p,d,q)(P,D,Q)_s, mostrado na Eq. 2.2.3.1. Esse modelo leva em consideração as componentes sazonais dos dados. (MORETTIN e TOLOI, 2018)

$$\phi_p(B)\Phi_P(B^S)(1 - B)^d(1 - B^S)^D Y_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t \quad (2.2.3.1)$$

Além dos termos já mostrado no modelo ARIMA (p,d,q), o modelo SARIMA(p,d,q)(P,D,Q)_s acrescenta o operador autorregressivo da parte sazonal de ordem P , $\Phi_P(B^S) = 1 - \Phi_1 B^S - \dots - \Phi_P B^{SP}$, o operador de médias móveis da parte sazonal de ordem Q , $\Theta_Q(B^S) = 1 - \Theta_1 B^S - \dots - \Theta_Q B^{SQ}$, e o operador diferença da parte sazonal $(1 - B^S)^D$, em que D é o número de diferenças sazonais para série ser estacionária, e S é o intervalo de tempo da sazonalidade. (MORETTIN e TOLOI, 2018)

2.2.4. Modelos SARIMAX (p,d,q)(P,D,Q)_sX

Todos os modelos já apresentados neste texto levam em consideração apenas informações referentes à variável que compõe a série temporal analisada. A metodologia de Box-Jenkins também possibilita a inserção de outras variáveis na modelagem. Neste caso, teríamos o modelo autoregressivo de médias móveis com regressores externos ou exógenos, chamado ARMAX, onde X é o vetor ou matriz das variáveis externas. Quando forem utilizadas diferenciações, tem-se o ARIMAX, modelo autoregressivo integrado de médias móveis com regressores externos ou

exógenos. Se houver sazonalidade, tem-se o SARIMAX $(p,d,q)(P,D,Q)_sX$, modelo sazonal autoregressivo integrado de médias móveis com regressores externos ou exógenos. O modelo SARIMAX insere na sua modelagem a combinação de dados defasados da série original e dados externos, como mostra a Eq. 2.2.4.1. (ARUNRAJ e AHRENS, 2015; VALER, 2016; U.S. CENSUS BUREAU, 2017).

$$Y_t = \Psi_0 + \Psi_1 X_{1,t} + \Psi_2 X_{2,t} + \dots + \Psi_k X_{k,t} + \frac{\theta_q(B)\Theta_Q(B^S)\varepsilon_t}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D} \quad (2.2.4.1)$$

em que: $X_{1,t}, X_{2,t}, \dots, X_{k,t}$ são observações de variáveis externas relacionadas à variável dependente Y_t e $\Psi_0, \Psi_1, \Psi_2, \dots, \Psi_k$, os coeficientes de regressão das variáveis externas. (ARUNRAJ e AHRENS, 2015).

2.3. Etapas da modelagem SARIMA

Nesta etapa são apresentadas e descritas as etapas empregadas na construção do modelo SARIMA $(p,d,q)(P,D,Q)_s$.

2.3.1. Especificação

Essa etapa tem como objetivo realizar uma análise exploratória dos dados para conhecer o comportamento da série, verificar a presença de tendência e sazonalidade, como também, avaliar a presença de estacionariedade.

Análise exploratória dos dados

De acordo com Morettin e Tolo (2018), uma análise exploratória da série permite verificar se os dados apresentam alguma tendência, variações sazonais e observações atípicas, além de outras propriedades estatísticas. Para Hyndman e Athanasopoulos (2018), a elaboração do gráfico em linhas de uma série temporal expõe algumas características que podem estar presentes na mesma. Além disso, para estes autores, a construção do gráfico em linhas a partir de sub-séries sazonais permite verificar claramente o padrão sazonal subjacente, como também mostrar as mudanças na sazonalidade ao longo do tempo.

Morettin e Tolo (2018) indicam a determinação de medidas descritivas, como: os valores máximo e mínimo, a tendência central (média e mediana) e a variabilidade (desvio-padrão e coeficiente de variabilidade), bem como assimetria e excesso de curtose, visto que todas estas medidas sintetizam informações da série temporal. Segundo Morettin e Tolo (2018) e Cowpertwait e Metcalfe (2009), a construção dos gráficos de caixas para verificar a distribuição dos dados também

auxilia na análise dos dados.

Além da análise exploratória, com aplicação de técnicas da estatística clássica, também é importante averiguar existência de sazonalidade e estacionariedade da série, para auxiliar na escolha de modelos de previsão mais adequados. A seguir são apresentados alguns testes utilizados para verificação da sazonalidade e estacionariedade.

Verificação da sazonalidade

Segundo Li et al. (2018), a Análise de Variância (ANOVA) permite identificar variação sazonal na demanda de energia elétrica, ou seja, analisar se há ou não diferenças estatísticas significativas entre os valores observados nos meses ao longo de um ano. De acordo com Barbetta et al. (2004), a ANOVA com blocos verifica se existem evidências de que a média do conjunto de observações, Y_{ij} , apresentado na Tab. 2.1, é diferente entre os g tratamentos (meses), considerando a presença de h blocos (anos).

A hipótese nula é que as médias são iguais entre todos os tratamentos: $H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_g$. A hipótese alternativa é que pelo menos uma média é diferente, isto é, $H_1: \mu_i \neq \mu_j$, para algum $i \neq j$ com $i=1,2,\dots,g$ (BARBETTA et al., 2004). A testagem é feita com a distribuição F de Snedecor-Fisher, sendo a estatística de teste (F_{calc}) calculada conforme descrito a seguir.

Tabela 2.1- Tabela cruzada dos dados

Blocos (h)	Tratamentos (g)			
	1	2	...	g
1	Y_{11}	Y_{21}	...	Y_{g1}
2	Y_{12}	Y_{22}	...	Y_{g2}
\vdots	\vdots	\vdots	\vdots	
h	Y_{1h}	Y_{2h}	...	Y_{gh}

Fonte: Barbetta et al., 2004.

Inicialmente calcula-se a soma de quadrados entre os tratamentos, conforme a Eq. 2.3.1.1, e a soma de quadrados entre os blocos, como mostra a Eq. 2.3.1.2, em que T , número total de observações, é dado por $T=gh$. (BARBETTA et al., 2004).

$$SQ_{Trat} = \sum_{i=1}^g \frac{Y_{i..}^2}{h} - \frac{Y_{..}^2}{T} \quad (2.3.1.1)$$

$$SQ_{Blocos} = \sum_{j=1}^h \frac{Y_{...j}^2}{g} - \frac{Y_{..}^2}{T} \quad (2.3.1.2)$$

Em seguida, calcula-se o quadrado médio entre os tratamentos, conforme Eq. 2.3.1.3, e entre os blocos, conforme Eq. 2.3.1.4.

$$QM_{Trat} = \frac{SQ_{Trat}}{g - 1} \quad (2.3.1.3)$$

$$QM_{Bloco} = \frac{SQ_{Bloco}}{h - 1} \quad (2.3.1.4)$$

O total da soma de quadrados é calculado conforme Eq. 2.3.1.5, sendo este resultado juntamente com os resultados encontrados nas Eq. 2.3.1.3 e 2.3.1.4 são utilizados para determinar o quadrado médio dos erros, conforme Eq. 2.3.1.6 e 2.3.1.7.

$$SQ_{Total} = \sum_{i=1}^g \sum_{j=1}^h Y_{ij}^2 - \frac{Y_{..}^2}{T} \quad (2.3.1.5)$$

$$SQ_{Erro} = SQ_{Tot} - SQ_{Trat} - SQ_{Bloco} \quad (2.3.1.6)$$

$$QM_{Erro} = \frac{SQ_{Erro}}{(g - 1)(h - 1)} \quad (2.3.1.7)$$

O p-valor é obtido pela distribuição F de Snedecor-Fisher, encontrando-se a probabilidade de ocorrência de resultados maiores ou iguais à estatística de teste (Eq. 2.3.1.8) com $(g-1)$ e $(g-1)(h-1)$ graus de liberdade.

$$F_{calc.} = \frac{QM_{Trat}}{QM_{Erro}} \quad (2.3.1.8)$$

Quando p-valor é menor que o nível de significância adotado, rejeita-se a hipótese nula, encontrando-se evidências de sazonalidade. Entretanto, para que seu resultado seja válido, é

necessário que sejam verificadas algumas condições deste modelo teórico, o que pode ser feito por uma análise de resíduos. Neste caso, deve-se verificar a presença de independência, normalidade e homocedasticidade. Quando estas propriedades não se verificam, um outro teste deve ser empregado, que segundo Siegel e Castellan Jr (2006), pode ser o teste não paramétrico de Friedman.

Para aplicação deste teste, primeiro atribui-se *ranks* (1,2,...,g) do menor para o maior valor, das observações, para cada bloco (*h*) da Tab. 2.1. Em seguida, somam-se os valores dos *ranks* de cada tratamento (R_j) e, por meio da Eq. 2.3.1.9, obtém-se uma estatística de teste que pode ser considerada como apresentando distribuição Quiquadrado com ($g-1$) graus de liberdade (MORETTIN e TOLOI, 2018; SIEGEL e CASTELLAN JR, 2006).

$$\chi^2_{calculado} = \left[\frac{12}{hg(g+1)} \sum_{j=1}^g R_j^2 \right] - 3h(g+1) \quad (2.3.1.9)$$

Esta distribuição é utilizada para encontrar o valor da probabilidade (p-valor), que comparado ao nível de significância permite decidir se a hipótese nula será rejeitada ou não. O teste de Friedman verifica se a hipótese nula de que as medianas do conjunto de observações Y_{ij} (Tabela 2.1) são iguais $H_0: M_{d1}=M_{d2}=M_{d3}=\dots=M_{dg}$, contra a hipótese alternativa de que ao menos duas observações apresentam medianas diferentes, ou seja, $H_1: M_{di} \neq M_{dj}$ para pelo menos duas condições ou grupos i e j (MORETTIN e TOLOI, 2018; SIEGEL e CASTELLAN JR, 2006).

Verificação da estacionariedade

Outra análise importante é a verificar se a série é ou não estacionária (MORETTIN e TOLOI, 2006). De acordo com Gujarati e Porter (2011), a análise da estacionariedade pode ser realizada a partir do teste de raiz unitária, em que a ausência desta permite encontrar evidências de que a série é estacionária. A verificação da presença de raiz unitária ocorre a partir de um processo estocástico de raiz unitária, mostrado da Eq. 2.3.1.10.

$$Y_t = \rho Y_{t-1} + \varepsilon_t, \text{ sendo } -1 \leq \rho \leq 1 \quad (2.3.1.10)$$

Quando $\rho=1$ existe uma raiz unitária e processo estocástico é identificado como não estacionário. Desta maneira, realiza-se a primeira diferença da Eq. 2.3.1.10, subtraindo Y_{t-1} em ambos os lados da igualdade apresentada na Eq. 2.3.1.10, e tem-se a Eq. 2.3.1.11, uma outra maneira de representar este mesmo processo (GUJARATI e PORTER, 2011).

$$\begin{aligned}
Y_t - Y_{t-1} &= \rho Y_{t-1} - Y_{t-1} + \varepsilon_t \\
Y_t - Y_{t-1} &= (\rho - 1)Y_{t-1} + \varepsilon_t \\
\nabla Y_t &= \delta Y_{t-1} + \varepsilon_t
\end{aligned} \tag{2.3.1.11}$$

A partir da Eq. 2.3.1.11, Dickey e Fuller (1979) propuseram testar a hipótese nula (H_0) de que $\delta=0$ ($\rho=1$), teste DF. Neste caso, tem-se uma raiz unitária e a série não é estacionária. Gujarati e Porter (2011) evidenciam que a verificação de estacionariedade de uma série consiste em averiguar o valor do coeficiente angular da Eq. 2.3.3.11. Quando a hipótese nula for rejeitada, não sendo possível concluir que $\delta = \rho - 1 = 0$, encontram-se evidências de que $\delta < 0$, representando uma série estacionária. Estes autores propuseram não utilizar a distribuição t de student para testar a significância deste coeficiente, mas a distribuição da estatística τ (tau). Usando simulações de Monte Carlo, criaram uma tabela com alguns valores tabelados (críticos) desta estatística, auxiliando nas decisões do teste. Se o valor da estatística tau (τ) exceder o valor crítico, rejeita-se a H_0 , ou seja, $\delta=0$ e neste caso encontram-se evidências de que a série é estacionária.

Dependendo da natureza do processo, ainda pode-se assumir outras duas composições para o teste DF. A primeira considerando o modelo apenas com a constante (β_1) (Eq 2.3.1.12), e a segunda considerando o modelo com constante (β_1) e com tendência ($\beta_2 t$), conforme Eq. 2.3.1.13.

$$\nabla Y_t = \beta_1 + \delta Y_{t-1} + \varepsilon_t \tag{2.3.1.12}$$

$$\nabla Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \varepsilon_t \tag{2.3.1.13}$$

Uma extensão deste teste, o Dickey-Fuller Aumentado (ADF) (Dickey e Fuller, 1981), considerou a presença da raiz unitária para modelos que apresentam erros autocorrelacionados. Segundo Gujarati e Porter (2011), o teste ADF faz uso da equação de regressão, conforme Eq. 2.3.1.14.

$$\nabla Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^m \phi_i \nabla Y_{t-i} + \varepsilon_t \tag{2.3.1.14}$$

em que: β_1 é o termo independente, também chamada de constante (*drift*), relacionado com o deslocamento da série, β_2 é o coeficiente de tendência, δ é o coeficiente de presença da raiz unitária e m é o número de atrasos utilizados. (GUJARATI e PORTER, 2011; VALER, 2016). Uma importante

observação é que a Eq. 2.3.1.14 inclui termos de diferenças, ∇Y_t , defasados empiricamente. A ideia principal é incluir termos suficientes para que os termos de erro destas equações sejam não correlacionados. (GUJARATI e PORTER, 2011). Além disso, dependendo da natureza do processo, de maneira similar ao teste DF, pode-se assumir as duas outras composições para o teste ADF. Segundo Morettin e Toloi (2018), esta etapa é importante, pois define a ordem de integração (d) do modelo ARIMA.

Para séries com sazonalidade, este mesmo teste pode ser utilizado. Neste caso, são realizadas duas análises: inicialmente para a parte não sazonal com o objetivo de determinar a ordem de integração (d), conforme feito anteriormente, e posteriormente para a parte sazonal com o objetivo de determinar a ordem de integração (D). Neste caso as diferenças são determinadas em função do período utilizado para analisar a sazonalidade. Por exemplo, se esta for mensal, a diferença seria entre o primeiro e o décimo terceiro termo.

Existem muitos testes para avaliar a estacionariedade de uma série temporal, tais como: Phillips-Perron (PP), Kwiatkowski-Phillips-Schmidt-Schin (KPSS), ADF-GLS e Zivot-Andrews (ZA), normalmente utilizados para análise da parte não sazonal, e Hylleberg-Engle-Granger-Yoo (HEGY) e Osborn-Chui-Smith-Birchenhall (OCSB), utilizados para analisar a estacionariedade da parte sazonal, entre outros. Mais detalhes sobre estes testes podem ser encontrados em Phillips e Perron (1988), Kwiatkowski et al. (1992), Elliot et al. (1996), Zivot Andrews (1992), Hylleberg et al. (1990) e Osborn et al. (1988).

O teste ADF é o mais utilizado na literatura pesquisada, principalmente para avaliar a estacionariedade da parte não sazonal, embora muitos autores, entre os quais Aylar et al (2019), tenham evidenciado o seu baixo poder. Para contornar este problema Silveira et al. (2017) sugerem, em caso de não rejeição da hipótese nula, complementar a análise com o teste KPSS.

2.3.2. Identificação

Nessa etapa determinam-se as ordens dos termos autorregressivo (p) e de médias móveis (q) da série temporal. Esses parâmetros são normalmente definidos através da interpretação gráfica conjunta das funções de autocorrelação e autocorrelação parcial. (MORETTIN e TOLOI, 2018; BUENO, 2011).

A função de autocorrelação (FAC) para um modelo MA(q) é mostrada na Eq. 2.3.2.1, e para o um modelo AR(p), na Eq. 2.3.2.2 (BUENO, 2011).

$$\rho_k = \frac{\theta_k + \theta_{k+1}\theta_1 + \theta_{k+2}\theta_2 + \dots + \theta_q\theta_{q-k}}{\sum_{k=0}^q \theta_k^2}, \quad k = 1, 2, \dots, q \quad (2.3.2.1)$$

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \dots + \phi_p\rho_{k-p}, \quad k = 1, 2, \dots, p \quad (2.3.2.2)$$

A FAC do modelo MA(q) torna-se zero a partir da defasagem q , ou seja, $\rho_k=0$ quando $k>q$, definindo a ordem do processo MA(q). Já a FAC do modelo AR(p) decai exponencialmente ao longo do tempo (BUENO, 2011). Este autor evidencia que este decaimento exponencial na FAC no modelo AR(p) revela que existe uma correlação implícita entre a série, Y_t , e a sua regressão, Y_{t-k} . Com a aplicação da função de autocorrelação parcial (FACP), é possível remover as correlações implícitas e manter apenas a correlação pura entre Y_t e Y_{t-k} . Para cada defasagem k utiliza-se a equação de regressão, Eq. 2.3.2.3.

$$Y_t = \phi_{k,1}Y_{t-1} + \phi_{k,2}Y_{t-2} + \dots + \phi_{k,k}Y_{t-k} + \varepsilon_t, \quad k = 1, 2, \dots, \quad (2.3.2.3)$$

Para $k=1$ considera-se uma regressão de Y_t contra Y_{t-1} e obtém-se $\phi_{1,1}$. Para $k=2$, tem-se uma regressão de Y_t contra Y_{t-1} e Y_{t-2} , obtendo-se $\phi_{2,1}$ e $\phi_{2,2}$, dos quais interessa apenas $\phi_{2,2}$. Ao considerar um modelo AR(p) genérico, serão encontrados coeficientes diferentes de zero até $\phi_{p,p}$ e coeficientes iguais a zero a partir de então, ou seja, ocorre um truncamento na ordem p do modelo AR(p). Devido a condição de inversibilidade, ao aplicar a FACP no modelo MA(q) esse resultado não ocorre, sendo que os coeficientes $\phi_{k,k}$, da Eq. 2.3.2.3, decaem exponencialmente (BUENO, 2011).

O comportamento gráfico da FAC e FACP para os modelos AR(p), MA(q) e ARMA(p,q) pode ser resumido através da Tab. 2.2. Já na Fig. 2.1 é apresentado alguns exemplos dos gráficos da FAC e FACP para os modelos AR(1) e MA(1).

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

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

Fonte: Souza e Camargo, 1996.

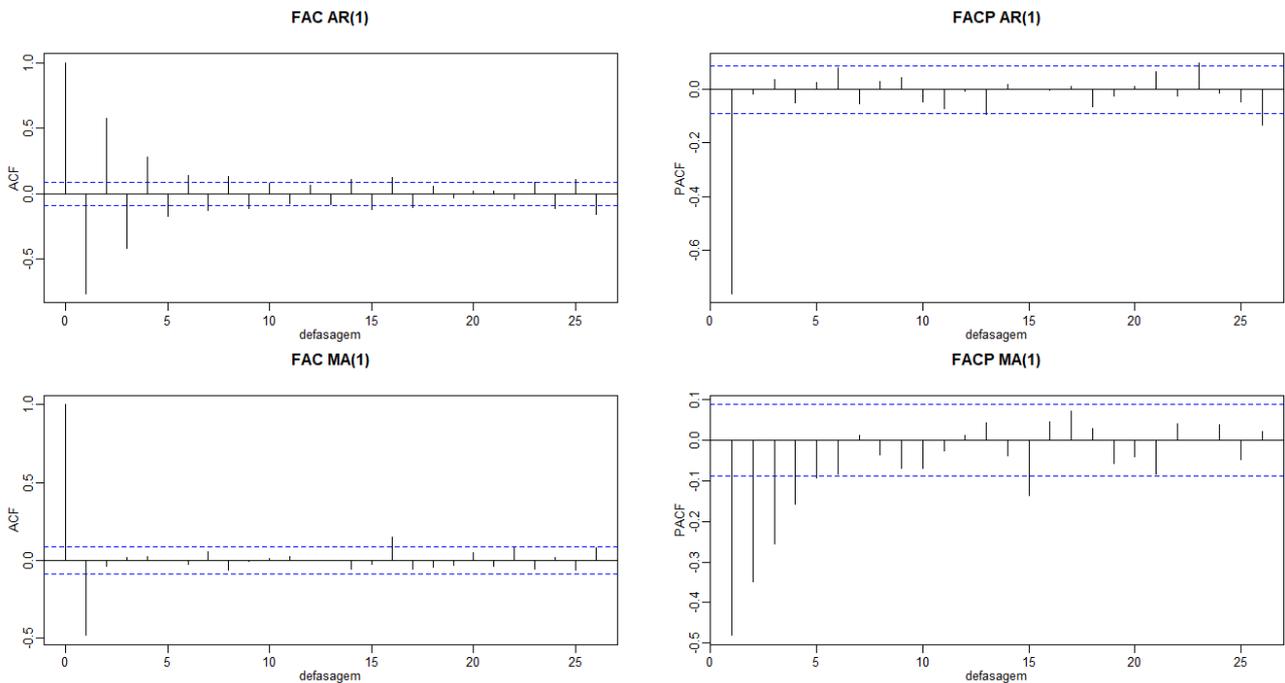


Figura 2.1 - Exemplos de gráficos da FAC e FACP para um modelo AR(1) e MA(1)

Na prática, entretanto, muitos autores, como Morettin e Toloí (2018), sugerem a definição das máximas ordens para p e q , estimando todas as combinações de modelos possíveis (modelos candidatos) e posteriormente, realizar um diagnóstico para identificar o modelo mais parcimonioso, que em geral, produz melhor previsão. Uma mesma série pode sugerir padrões alternativos de modelagem, devido à dificuldade de deduzir graficamente qual é o padrão da função que está gerando a série.

2.3.3. Estimação dos parâmetros

A estimação dos parâmetros pode ser obtida por diferentes métodos: mínimos quadrados ordinários, mínimos quadrados generalizados, máxima verossimilhança condicional ou máxima verossimilhança exata, entre outros. De acordo com Bueno (2011), as estimativas obtidas por esses métodos são todas consistentes, diferentemente da eficiência, que é maior no método de máxima verossimilhança exata, o mais citado na literatura na modelagem de séries temporais.

De acordo com Morettin e Toloí (2018), dependendo da escolha dos valores iniciais para estimação dos parâmetros empregados na função verossimilhança, pode ser utilizada máxima verossimilhança condicional ou máxima verossimilhança exata. Na condicional, os valores iniciais desconhecidos são substituídos por valores que supõe serem razoáveis, enquanto que na exata, os valores iniciais são estimados utilizando um procedimento denominado “*backforecasting*”. Segundo

Thomaz (2019), embora a máxima verossimilhança condicional seja mais fácil de ser empregada, ambos os métodos podem ser utilizados para amostras grandes e que apresentem estacionariedade e ergodicidade. Os cálculos da função de máxima verossimilhança exata são bastante complexos e podem ser encontrados com mais detalhes em Hamilton (1993) e Morettin e Toloí (2018).

Segundo Bueno (2011), para realizar a estimação por máxima verossimilhança, assume-se que os dados tenham uma distribuição normal. Entretanto, os seus resultados também são satisfatórios, mesmo que esta condição não seja satisfeita.

De acordo com Bueno (2011), o melhor modelo será o mais parcimonioso, ou seja, aquele que apresenta um menor número de parâmetros no modelo, e com isso, seja capaz de gerar estimativas com menores imprecisões do que modelos que apresentem mais parâmetros inseridos. Esta propriedade pode ser verificada pelos critérios de informação.

Ainda, segundo Morettin e Toloí (2018), após a estimação de todos os parâmetros de um modelo, deve-se verificar se este representa os dados adequadamente, existindo necessidade de realização de uma análise de diagnóstico.

Critério de Informação

Segundo Cryer e Chan (2008), o Critério de Informação Akaike (AIC), determinado conforme Eq. 2.3.3.1, e o Critério de Informação Bayesiana (BIC), determinado conforme Eq. 2.3.3.2, podem ser utilizados para selecionar aquele que seria o mais parcimonioso, evitando selecionar modelos com muitos parâmetros. Esses critérios empregam a estimativa do log da máxima verossimilhança, penalizando os ajustes dos modelos pelo número de parâmetros estimados. São eles:

$$AIC = -2 \log(L) + 2(r) \quad (2.3.3.1)$$

$$BIC = -2 \log(L) + r \log(T) \quad (2.3.3.2)$$

onde: $\log(L)$ é o logaritmo da função máxima verossimilhança. O termo $2r$ tem a função de penalidade, sendo r o número total de parâmetros, ou seja, $r=p+q+1$, se o modelo estimado tiver constante e, $r=p+q$ se não houver constante. Já T é o tamanho da amostra analisada. O modelo a ser escolhido deve ser aquele que apresente os menores valores para o(s) critério(s) considerado(s) na análise (CRYER e CHAN, 2008; BUENO, 2011).

Além destes, outros critérios têm sido utilizados com esta finalidade, como o Critério de Informação de Akaike Corrigido (AICc), determinado conforme Eq. 2.3.3.3 (THOMAZ, 2019).

$$AICc = AIC + \frac{2(p+q)(p+q+1)}{T-(p+q)-1} \quad (2.3.3.3)$$

Mais detalhes sobre estes critérios de informação podem ser encontrados em Akaike (1973), Schwarz (1978) e Hurvitch e Tsai (1991).

Análise dos resíduos

Após a determinação da parcimonialidade dos modelos candidatos, é necessário avaliar a sua adequação aos dados da série, o que pode ser feito por uma análise de resíduos, sendo esperado que estes sejam ruído branco. Caso o modelo estimado apresente resíduos que não sejam caracterizados como ruído branco, ainda há informações nos dados não captadas pelo modelo, ou seja, as previsões geradas por ele serão pobres. Dessa maneira, é aconselhável descartar o modelo estimado e testar outras possibilidades (BUENO, 2011).

A primeira propriedade analisada nos resíduos é a presença de autocorrelação, que pode ser verificada pelo teste de Ljung-Box (LB), proposto em Ljung e Box (1978). A hipótese nula (H_0) é de não existir evidências de que os resíduos sejam autocorrelacionados, enquanto a hipótese alternativa (H_1) é de ter evidências que os resíduos sejam autocorrelacionados. A estatística deste teste utilizada nesta verificação, apresentada na Eq. 2.3.3.4, converge para uma distribuição χ^2 com $k-p-q$ graus de liberdade (CRYER e CHAN, 2008; BUENO, 2011; GUJARATI e PORTER, 2011).

$$LB = T(T+2) \left(\frac{\hat{\rho}_1^2}{T-1} + \frac{\hat{\rho}_2^2}{T-2} + \dots + \frac{\hat{\rho}_k^2}{T-k} \right) \quad (2.3.3.4)$$

em que: T é o tamanho da série, $\hat{\rho}_j$ é uma estimativa de autocorrelação com j defasagens ($j = 1, 2, \dots, k$). De acordo com Morettin e Toloí (2018), em geral basta utilizar as 10 ou 15 primeiras defasagens. Entretanto, de acordo com estes autores, recentemente tem havido críticas em relação a esta medida por ser considerado que ela avalia a correlação e não a dependência.

O teste Jarque-Bera (JB), proposto em Jarque e Bera (1986), pode ser utilizado para verificar a normalidade dos resíduos. Sua estatística de teste está apresentada na Eq. 2.3.3.5, em que A é o coeficiente de assimetria e K o coeficiente de curtose. A hipótese nula (H_0) deste teste assume que os resíduos seguem uma distribuição Normal e a estatística de teste utilizada segue uma distribuição Quiquadrado (χ^2) com 2 graus de liberdade. (CRYER e CHAN, 2008; BUENO, 2011; GUJARATI e PORTER, 2011).

$$JB = T \left[\frac{A^2}{6} + \frac{(K-3)^2}{24} \right] \quad (2.3.3.5)$$

Além do teste de Jarque-Bera, outros testes como o de Shapiro-Wilk e o de Kolmogorov-Smirnov Lilliefors podem ser utilizados para avaliar normalidade. Métodos gráficos, como o gráfico estatístico Quantil-Quantil (QQ), podem complementar a análise de normalidade dos resíduos (CRYER e CHAN, 2008). A exigência de normalidade é necessária se for utilizada alguma técnica que suponha normalidade, facilitando a posterior análise inferencial.

O teste ARCH verifica a homocedasticidade dos resíduos de um modelo, sendo desejável que estes sejam homocedásticos (BUENO, 2011). A Eq. 2.3.3.6 mostra um modelo de regressão linear de k variáveis, dependendo da informação disponível no tempo $(t-1)$, em que: $\varepsilon_t \sim N[0, (\alpha_0 + \alpha_1 \varepsilon_{t-1}^2)]$. (GUJARATI e PORTER, 2011).

$$Y_t = \beta_0 + \beta_1 X_{1t} + \dots + \beta_k X_{kt} + \varepsilon_t \quad (2.3.3.6)$$

O termo ε_t é normalmente distribuído com média zero e variância como mostra a Eq. 2.3.3.7, ou seja, seguindo um processo do tipo ARCH(1) para este caso analisado. Os termos α_0 e α_1 , desta equação, representa os coeficiente associados aos erros. (GUJARATI e PORTER, 2011)

$$var(\varepsilon_t) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \quad (2.3.3.7)$$

Percebe-se na Eq. 2.3.3.7, que a variância de ε no período t é dependente do quadrado do distúrbio 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 Eq. 2.3.3.8. (GUJARATI e PORTER, 2011).

$$var(\varepsilon_t) = \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_p \varepsilon_{t-p}^2 \quad (2.3.3.8)$$

Segundo Gujarati e Porter (2011), nesta análise, se não houver nenhuma autocorrelação na variância de erros na Eq. 2.3.3.7, a hipótese nula será $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_p = 0$, sendo o caso em que $var(\varepsilon_t) = \alpha_0$, e não ocorre o efeito ARCH. Desta maneira, diz-se que os resíduos são homocedásticos. Como hipótese alternativa (H_1) tem-se que os resíduos são heteroscedásticos.

Como não é possível obter diretamente o valor de σ_t^2 , pode-se regredir a Eq. 2.3.3.8, conforme

Eq. 2.3.3.9, e testar a hipótese nula.

$$\hat{\varepsilon}_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1 \hat{\varepsilon}_{t-1}^2 + \hat{\alpha}_2 \hat{\varepsilon}_{t-2}^2 + \dots + \hat{\alpha}_p \hat{\varepsilon}_{t-p}^2 \quad (2.3.3.9)$$

em que: $\hat{\varepsilon}_t$ é uma estimativa dos resíduos obtidos por mínimos quadrados ordinários obtidos em um modelo de regressão (Eq. 2.3.3.6). Desta maneira, pode-se testar H_0 através da distribuição Qui-quadrado, com p graus de liberdade. A estatística de teste é mostrada na Eq. 2.3.3.10.

$$ALM_p = TR^2 \quad (2.3.3.10)$$

em que: T é o número total de observações, p é a quantidade de defasagens consideradas na Eq. 2.3.3.9.

2.3.4. Validação

Nesta etapa são apresentados os indicadores de qualidade, que são medidas de acurácia, utilizadas para avaliar o desempenho da modelagem realizada para previsão. Segundo Scheidt et al. (2020), os três indicadores mais utilizados na validação de um modelo são: Erro Absoluto Médio (MAE – *Mean Absolute Error*), Erro Médio Percentual Absoluto (MAPE – *Mean Absolute Percentage Error*) e Raiz do Erro Quadrático Médio (RMSE – *Root Mean Squared Error*).

O MAE, apresentado na Eq. 2.3.4.1, é a soma em módulo da diferença entre o valor observado (Y_t) e previsto (\hat{Y}_t), para cada intervalo de tempo analisado, dividido pelo número total de observações no período de validação da série (H). Este indicador de qualidade mostra a distância entre os valores observados e previstos, sendo de fácil interpretação visto que trata de valores absolutos, apresentando a mesma unidade da variável analisada (LEVENBACH, 2017; SCHEIDT et al, 2020).

$$MAE = \frac{1}{H} \sum_{t=1}^H |Y_t - \hat{Y}_t| \quad (2.3.4.1)$$

O MAPE, apresentado na Eq. 2.3.4.2, representa a média da diferença absoluta entre os valores previstos e observados, expressa em porcentagem dos valores observados. Este indicador, tem a vantagem de ser adimensional. (LEVENBACH, 2017; SCHEIDT et al, 2020).

$$MAPE = \left(\frac{1}{H} \sum_{t=1}^H \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \right) \cdot 100 \quad (2.3.4.2)$$

O RMSE, apresentado na Eq. 2.3.4.3, provém do Erro Quadrático Médio (MSE – Mean Squared Error), sendo determinado pela raiz quadrada da soma das diferenças entre o valor observado e o valor previsto elevadas ao quadrado e dividida pelo número de períodos analisados. Os resultados obtidos a partir do RMSE utilizam os quadrados dos resíduos, razão pela qual os maiores resíduos têm um peso maior em seu resultado. Vários autores, entre os quais Morettin e Toloí (2018), o consideram a melhor medida de acurácia. Já Levenbach (2017) ressalta que esse indicador pode dificultar a interpretação dos dados, dependendo do tipo de série analisada. Um estudo feito por Scheidt et al. (2020) mostra que esse indicador é o mais utilizado em estudos envolvendo o setor elétrico, entretanto eles ressaltam que é adequado a sua utilização de outras medidas para complementar a análise.

$$RMSE = \left[\frac{\sum_{t=1}^H (Y_t - \hat{Y}_t)^2}{H} \right]^{\frac{1}{2}} \quad (2.3.4.3)$$

Além desses indicadores de qualidades, segundo LEVENBACH (2017), pode ser utilizada mais uma medida da capacidade preditiva do modelo. Trata-se de aplicar uma técnica desenvolvida por Theil (1966) que tem como objetivo comparar se os resultados da previsão, apresentadas pelo modelo encontrado, são melhores do que considerar o método ingênuo para previsão. Este indicador é calculado pela Eq. 2.3.4.4.

$$U = \sqrt{\frac{\sum_{t=1}^{H-1} \left(\frac{\hat{Y}_{t+1} - Y_{t+1}}{Y_t} \right)^2}{\sum_{t=1}^{H-1} \left(\frac{Y_{t+1} - Y_t}{Y_t} \right)^2}} \quad (2.3.4.4)$$

Espera-se que o resultado deste indicador seja inferior a 1, permitindo-se afirmar que as previsões do modelo encontrado são melhores que as previsões ingênuas (quando a previsão para variável Y no instante t é igual ao valor de Y no instante $t-1$). Caso seja igual a 1, considera-se que as previsões não são nem melhores, nem piores com as obtidas pelas previsões ingênuas. Já nas situações

em que for maior que 1, este modelo é descartado, pois suas previsões são piores que as obtidas pelo método ingênuo.

3. METODOLOGIA

Neste capítulo apresenta-se os procedimentos metodológicos utilizados na aplicação da modelagem SARIMA e SARIMAX. Para a realização deste trabalho, foram utilizadas séries temporais com dados mensais, do período de janeiro de 2004 a dezembro de 2019, totalizando 192 observações em cada uma das três classes consumidoras: residencial, comercial e industrial. Na Tab. 3.1 são apresentadas as séries temporais empregadas e suas respectivas bases de dados.

Tabela 3.1 - Dados e fontes

Variável	Fonte
Demanda mensal de energia elétrica da classe residencial (Dem_{res})	EPE (2020)
Demanda mensal de energia elétrica da classe comercial (Dem_{com})	EPE (2020)
Demanda mensal de energia elétrica da classe industrial (Dem_{ind})	EPE (2020)
Tarifa de energia elétrica residencial (Tar_{res})	ANEEL (2020)
Tarifa de energia elétrica comercial (Tar_{com})	ANEEL (2020)
Tarifa de energia elétrica industrial (Tar_{ind})	ANEEL (2020)
Número de consumidores residencial (NC_{res})	EPE (2020)
Número de consumidores comercial (NC_{com})	EPE (2020)
Número de consumidores industrial (NC_{ind})	EPE (2020)
IPCA classe residencial	IPEA (2020)
IPCA classe comercial	IPEA (2020)
IPCA classe industrial	IPEA (2020)
Temperatura (Temp)	INMET (2020)

Fonte: Elaborada pela autora.

Todos os procedimentos descritos a seguir foram realizados com o auxílio dos *softwares* Gnu Regression, Econometrics and Time-Series Library (Gretl, 2020), Rstudio (R CORE TEAM, 2020) e da planilha Excel. Nas análises inferenciais foram utilizados 0,01 e 0,05 como nível de significância.

3.1. Procedimentos metodológicos da modelagem SARIMA

Nesta etapa foram desenvolvidos modelos empregando apenas a série temporal da demanda mensal de energia elétrica para as classes consumidoras do Rio Grande do Sul: residencial (Dem_{res}), comercial (Dem_{com}) e industrial (Dem_{ind}). A Fig. 3.1 apresenta um esquema das etapas desenvolvidas

para obtenção do modelo SARIMA.

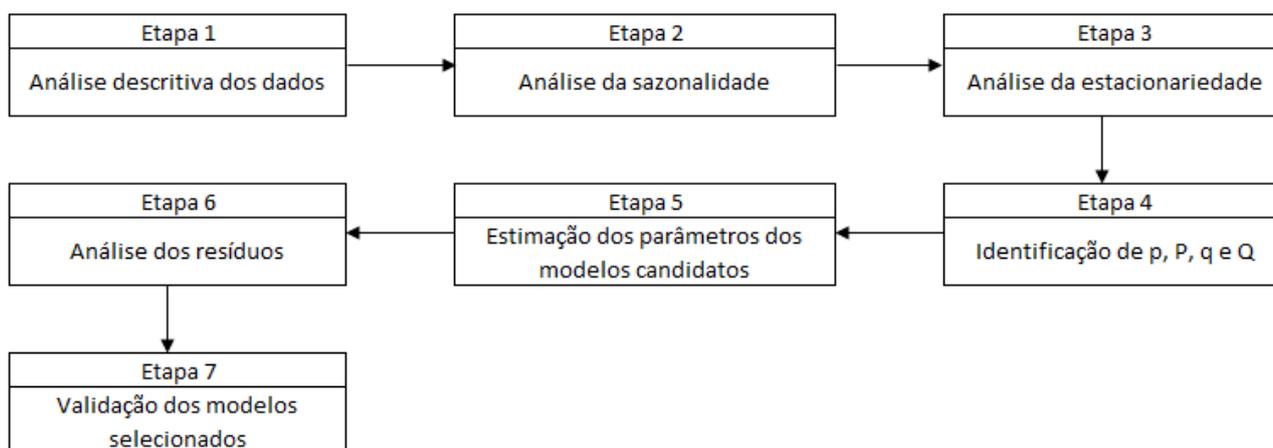


Figura 3.1 - Esquema das etapas metodológicas para modelagem SARIMA

Na etapa 1, realizou-se a análise exploratória dos dados da demanda mensal de energia elétrica para cada uma das três classes consumidoras. Nesta etapa foram elaborados gráficos em linha e gráficos das subséries sazonais e estas análises foram complementadas com o levantamento das medidas descritivas: valores máximo e mínimo, desvio padrão, média, coeficiente de assimetria, coeficiente de curtose e coeficiente de variação, bem como a construção do gráfico *box plot* por mês e por ano, mostrando a distribuição dos mesmos.

A etapa 2 compreendeu a fase de verificação da presença de sazonalidade das séries estudadas. Inicialmente foi empregado o teste ANOVA com blocos, e em seguida foi verificada a autocorrelação, a normalidade e a heterocedasticidade dos resíduos. Quando uma destas três condições não foi satisfeitas, aplicou-se o teste de Friedman.

Antes de desenvolver a etapa 3 foi analisada a variabilidade das séries e aplicada a transformação logarítmica para diminuí-la. Na sequência, realizou-se o janelamento dos dados, separando os primeiros 180 para fase de ajuste e 12 últimos para fase de validação. Em seguida, foi verificada a estacionariedade das séries através do teste ADF, o que foi feito separadamente para a parte não sazonal e para a parte sazonal.

Na etapa 4 identificou-se a ordem dos parâmetros autoregressivos 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, utilizou-se a função *auto.arima* disponível no *software* Rstudio. Com o resultado apresentado por esta função, optou-se por atribuir os valores 0, 1, 2 e 3 a cada um dos parâmetros p, q, P e Q, considerando todas as

combinações possíveis entre estes. Resultaram 256 modelos.

Utilizando o método da máxima verossimilhança, na etapa 5, foi feita a estimação dos parâmetros que compõem os modelos. Na sequência, estes foram hierarquizados, pela parcimonia através dos valores do critério de informação Bayesiano (BIC). Na sequência, foram selecionados os 13 modelos mais parcimoniosos, que representam 5% do conjunto total.

Já na etapa 6 foi realizado o diagnóstico dos resíduos obtidos com os 13 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 7, realizou-se a validação dos modelos. Considerando os 12 últimos valores da amostra foram analisados os indicadores de qualidade: MAE, MAPE, RMSE e U de Theil. Com a escolha do modelo que melhor realiza previsões, foram construídos intervalos de confiança, representados também de forma gráfica.

3.2. Procedimentos metodológicos para o refinamento do modelo: SARIMAX

Nesta etapa foram desenvolvidos outros modelos, pois além da demanda mensal de energia elétrica das três principais classes consumidoras do estado do Rio Grande do Sul, também são inseridas variáveis externas. Tem-se como objetivo, ao acrescentar as variáveis explicativas, obter modelos com melhor desempenho que os encontrados na etapa anterior, modelagem SARIMA. A Fig. 3.2 apresenta um esquema das etapas desenvolvidas para obtenção do modelo SARIMAX.

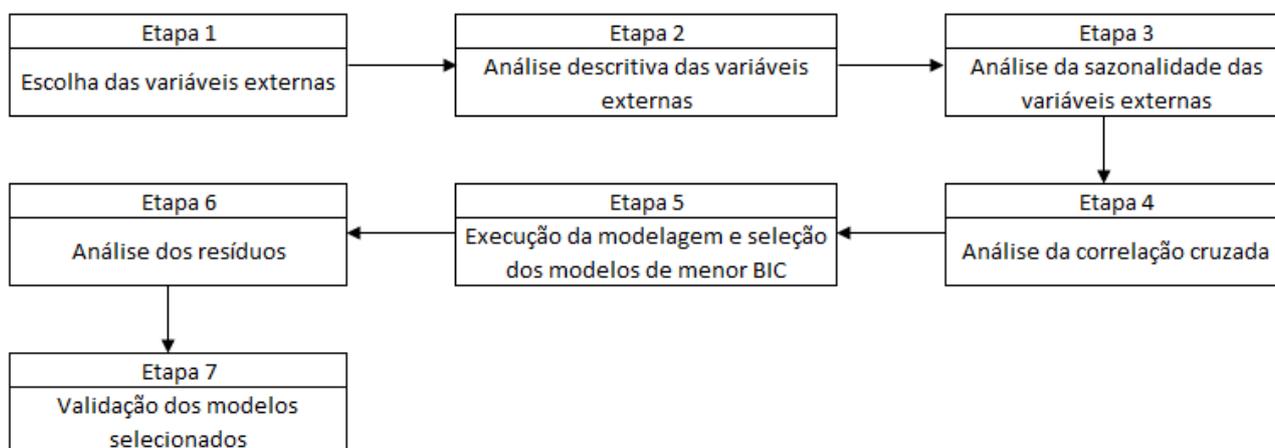


Figura 3.2 – Esquema das etapas metodológicas para modelagem SARIMAX

A primeira etapa consiste na escolha das variáveis externas, sendo avaliada a possibilidade de inserção de três variáveis. A primeira variável foi o número de consumidores, pois o seu aumento deve acarretar o aumento da demanda de energia elétrica, conforme já mostrado em Pereira (2015). A segunda variável foi a tarifa, já utilizada em outros estudos, como em Silveira (2017) e Nunes (2018), que a relacionam à demanda de energia elétrica. Cabe ressaltar que a variável Índice Nacional de Preços ao Consumidor Amplo (IPCA), presente na Tab. 3.1, foi empregada para deflacionar a tarifa, ou seja, remover os efeitos causados pela inflação. A terceira variável é a temperatura média do estado do Rio Grande do Sul, sendo comum aos modelos das três classes. Esta variável também já foi empregada em estudos de demanda de energia elétrica, como em Pereira (2015).

A etapa 2 consistiu em uma análise descritiva dos dados, enquanto na etapa 3 foi verificada a presença de sazonalidade das séries, empregando os mesmos procedimentos e critérios já descritos no item anterior.

Na etapa 4 foi feita uma análise de correlação cruzada cujo objetivo é verificar a relação entre a demanda mensal de energia elétrica e possíveis variáveis externas e suas defasagens, eliminando o efeito aleatório. Esta análise foi realizada com os valores observados de cada variável, considerando parte não sazonal e variações sazonais, da mesma maneira que no teste ADF. Nesta análise foi identificada a maior defasagem significativa e realizadas modelagens sucessivas deste valor máximo até 1, sendo identificado o modelo mais parcimonioso pelo BIC. Com isto foi possível definir a quantidade de defasagens a ser utilizada na modelagem posterior.

Na etapa 5 foram executados modelos candidatos, considerando a demanda mensal de energia elétrica e cada uma das variáveis externas e suas defasagens identificadas na etapa anterior, sendo selecionada a defasagem ideal pelo critério de informação Bayesiano (BIC). Com estas defasagens definidas, foram analisadas todas as combinações possíveis entre as variáveis externas e a demanda mensal de energia elétrica. Novamente pelo BIC, foi selecionado o modelo mais parcimonioso, seja com apenas uma única variável externa ou com mais de uma variável.

Na etapa 6 foram analisados os resíduos do modelo mais parcimonioso, e por fim, na etapa 7, fez-se a validação destes modelos. Nestas etapas foram utilizados os mesmos critérios e técnicas aplicados na modelagem SARIMA

O procedimento acima descrito foi realizado separadamente em cada uma das três classes consumidoras: residencial, comercial e industrial.

4. RESULTADOS E DISCUSSÕES

Neste capítulo são apresentados resultados e discussões da análise de dados para as três classes consumidoras. Na primeira seção são mostrados os resultados da modelagem SARIMA e na segunda, o refinamento do modelo com a inserção de variáveis externas: modelagem SARIMAX. Já na última seção é feita uma discussão sobre o desempenho entre os modelos desenvolvidos.

4.1. Modelagem SARIMA

As figuras 4.1, 4.2 e 4.3 mostram os gráficos das séries temporais da demanda mensal de energia elétrica para as classes residencial, comercial e industrial, respectivamente, para o estado do Rio Grande do Sul, no período de 2004 até 2019. Em todas as classes analisadas notam-se picos em determinados períodos do ano, sugerindo a presença de sazonalidade nas séries.

As classes residencial e comercial, Fig. 4.1 e Fig. 4.2, respectivamente, apresentam um padrão de crescimento da demanda maior (tendência) ao longo dos anos, se comparadas à demanda da classe industrial, Fig. 4.3. Além disso, estas duas séries também apresentam um aumento perceptível da sua variabilidade. Outro ponto comum entre essas duas séries é a presença de um pico no ano de 2014, seguido de um decaimento no período de 2015 a 2016 e, posteriormente, uma nova tendência de crescimento nos anos seguintes apenas na classe residencial.

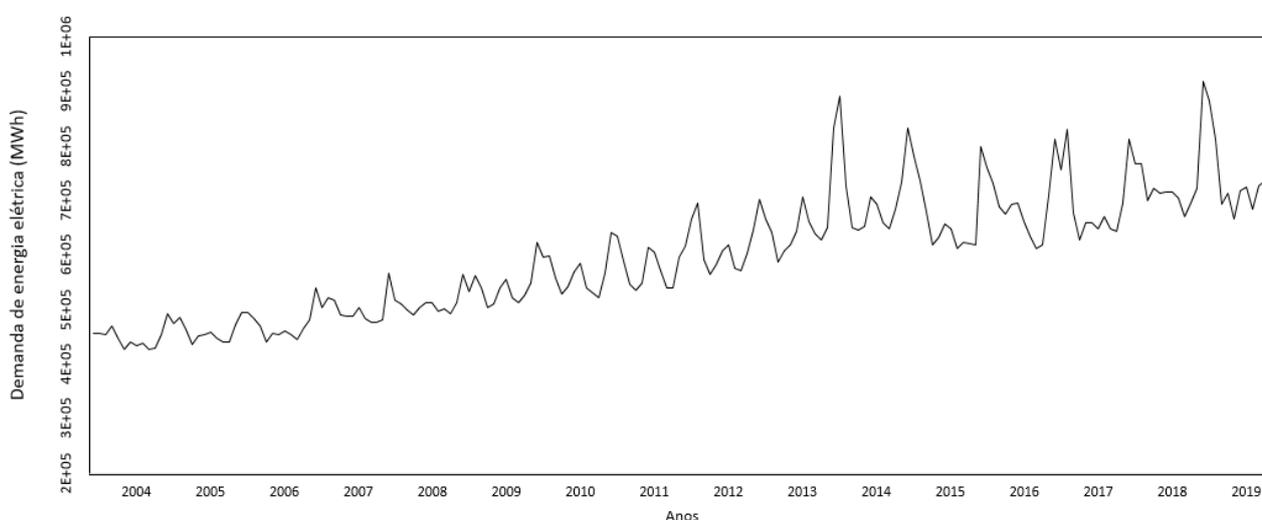


Figura 4.1 – Demanda mensal de energia elétrica da classe residencial/RS, 2004 - 19

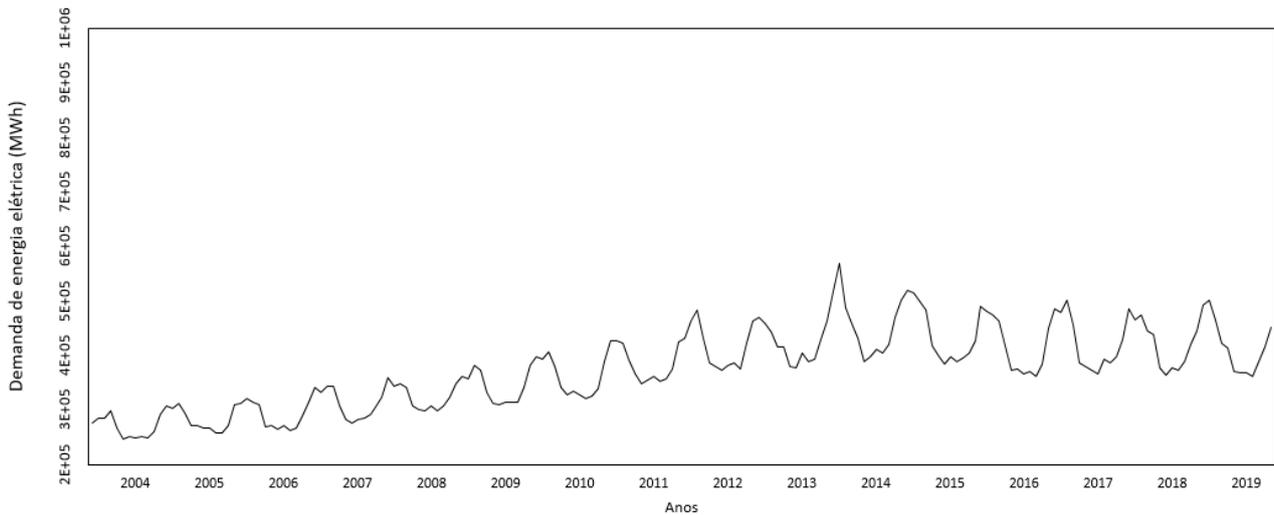


Figura 4.2 - Demanda mensal de energia elétrica da classe comercial/RS, 2004 - 19

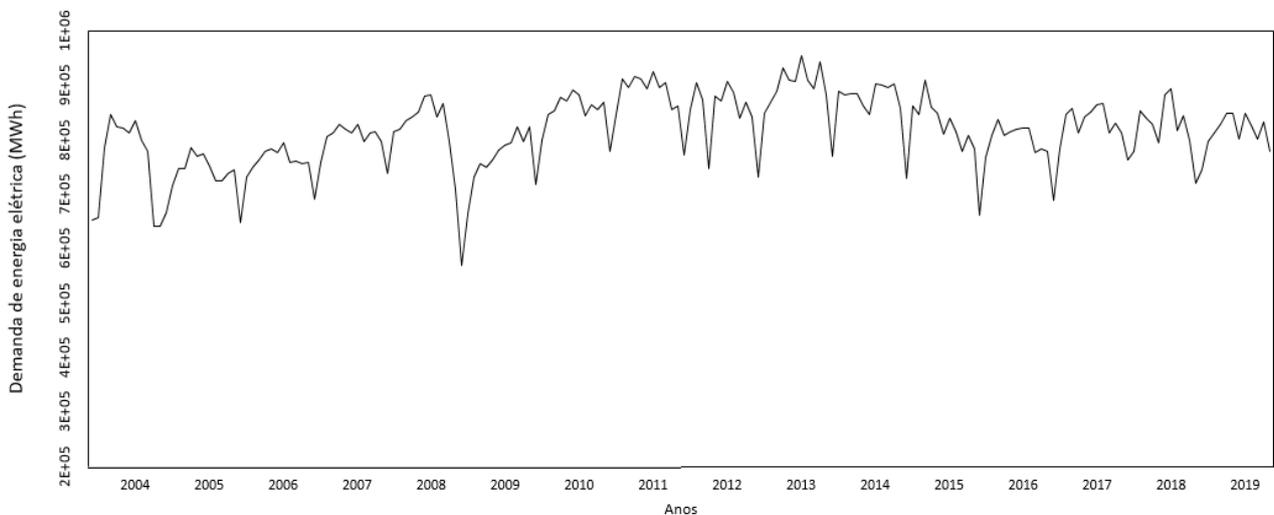


Figura 4.3 - Demanda mensal de energia elétrica da classe industrial/RS, 2004 - 19

Embora a série da demanda industrial aparentemente apresente características sazonais, esta não acompanhou o mesmo comportamento das outras duas classes. Percebe-se, na Fig. 4.3, que um valor mínimo da série é apontado em 2009, ano em que ocorreu a crise de abastecimento de energia no país, conforme já destacado em outros trabalhos: Silveira (2017) e Nunes (2019).

Os gráficos das subséries sazonais, que mostram a tendência central mensal das séries ao longo dos anos para as três classes consumidoras, são mostradas nas figuras 4.4, 4.5 e 4.6 correspondendo, respectivamente, às classes residencial, comercial e industrial. Pode-se verificar que as séries temporais analisadas indicam a presença da componente sazonal, uma vez que, para cada mês, a tendência central e a variabilidade diferem.

As classes residencial (Fig. 4.4) e comercial (Fig. 4.5) apresentam janeiro, fevereiro e março

como os meses com maior média de demanda, o mesmo não acontecendo na classe industrial (Fig. 4.6). As demandas médias com valores intermediários, bem como os menores valores médios de demanda, variam de acordo com cada classe.

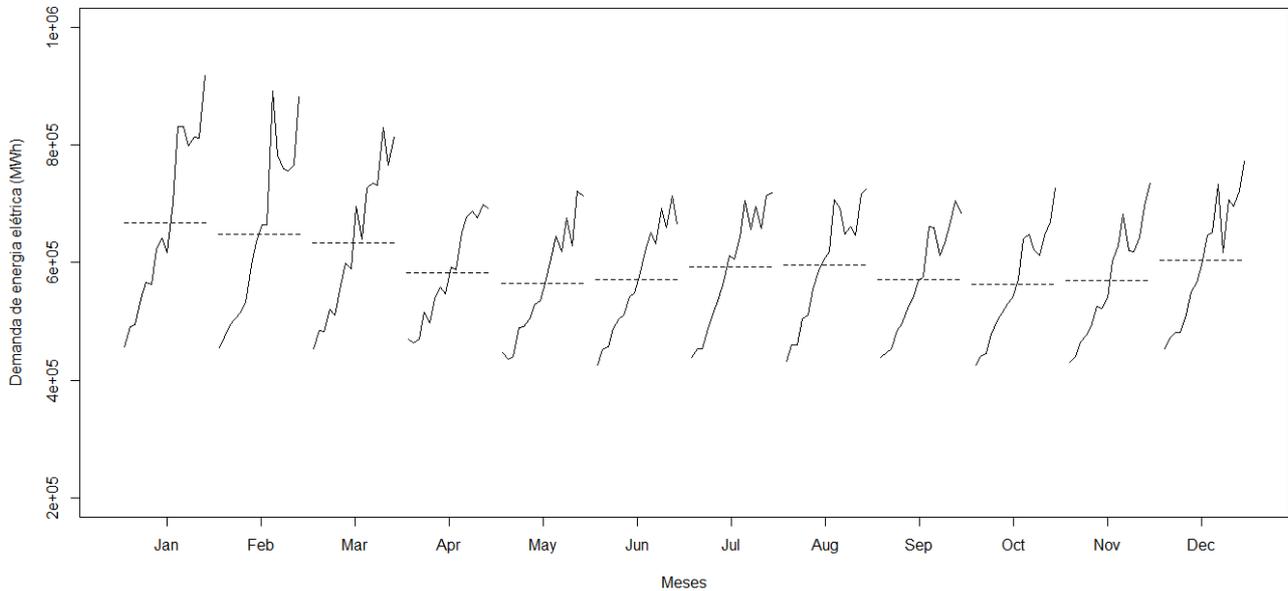


Figura 4.4 - Representação do valor da tendência central para cada mês ao longo dos anos da demanda mensal de energia elétrica para a classe residencial/RS, 2014 - 19

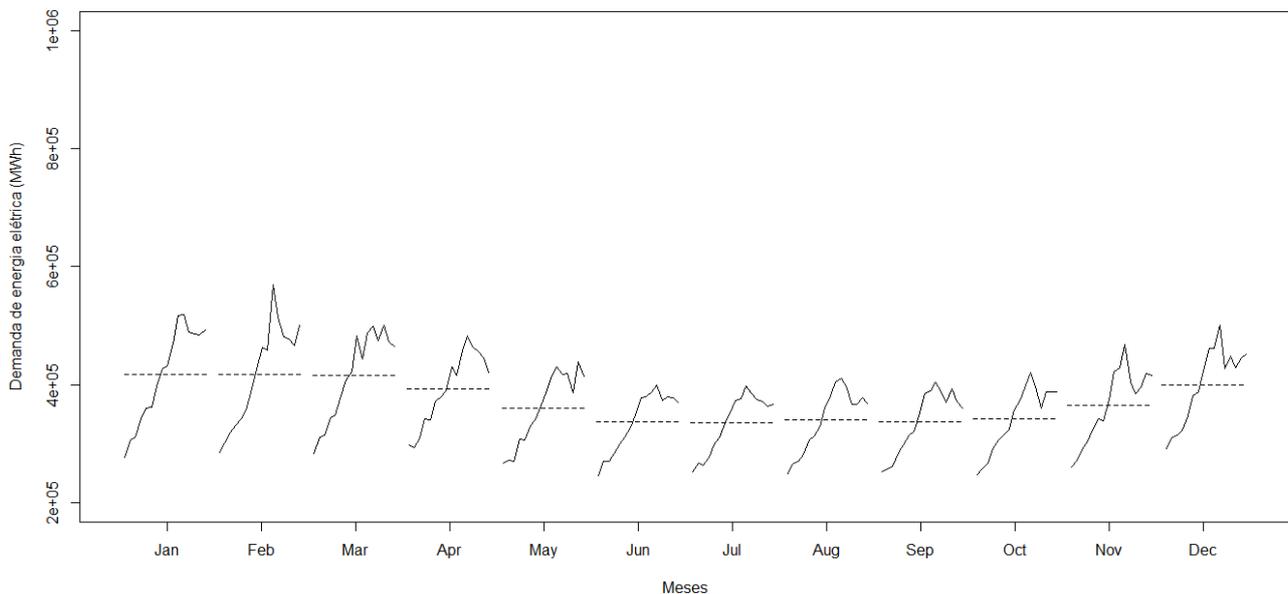


Figura 4.5 - Representação do valor da tendência central para cada mês ao longo dos anos da demanda mensal de energia elétrica para a classe comercial/RS, 2004 - 19

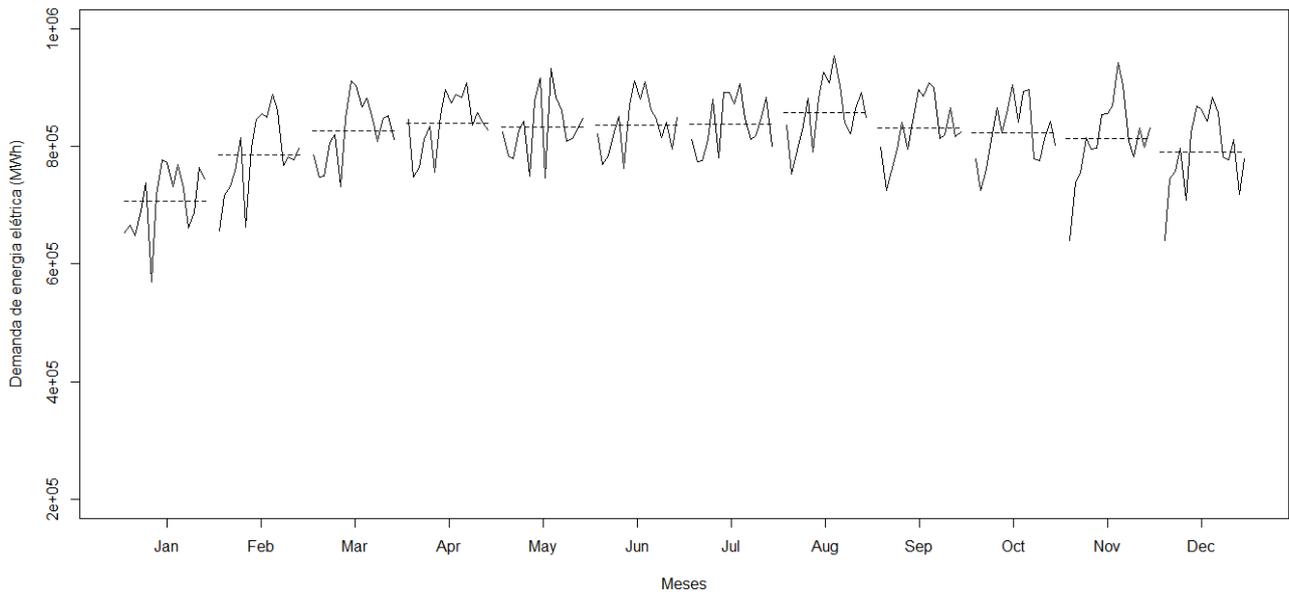


Figura 4.6 - Representação do valor da tendência central para cada mês ao longo dos anos da demanda mensal de energia elétrica para a classe industrial/RS, 2004 - 19

Essas análises gráficas são complementadas com os valores das medidas descritivas, apresentadas na Tab. 4.1. Em relação à tendência central, observa-se que ela é mais elevada na classe industrial e mais baixa na classe comercial, o mesmo acontecendo em relação aos valores máximo e mínimo. Pode-se verificar também que a demanda na classe residencial apresenta maior amplitude total em relação às outras classes, o mesmo acontecendo com o desvio-padrão. Ainda, a amostra da demanda na classe industrial apresenta menor coeficiente de variação em relação às outras classes, ou seja, suas observações são mais homogêneas.

Os coeficientes de assimetria indicam que os dados das classes residencial e comercial podem ser considerados simétricos. Já a classe industrial aponta para uma assimetria moderada à esquerda (negativa). Os coeficientes de curtose apontam para uma curva platicúrtica nas classes residencial e comercial, e leptocúrtica para classe industrial.

Tabela 4.1 - Medidas resumo para as três classes consumidoras da demanda mensal de energia elétrica/RS, 2004 - 19

Parâmetro	Residencial	Comercial	Industrial
Valor máximo (MWh)	917.352,0	569.657,0	954.097,0
Valor mínimo (MWh)	425.987,0	245.687,0	568.855,0
Média (MWh)	597.089,0	372.021,0	815.027,0
Mediana (MWh)	597.478,0	373.587,0	820.164,0
Desvio padrão (MWh)	110.228,9	70.390,18	66.506,0
Coef. de variação (%)	18,5	18,9	8,2
Coef. de assimetria	0,4	0,2	-0,7
Coef. de curtose	-0,5	-0,7	0,6

Fonte: Elaborada pela autora.

As figuras 4.7, 4.8 e 4.9 apresentam o gráfico de caixa para cada ano da demanda mensal de energia elétrica residencial, comercial e industrial, respectivamente.

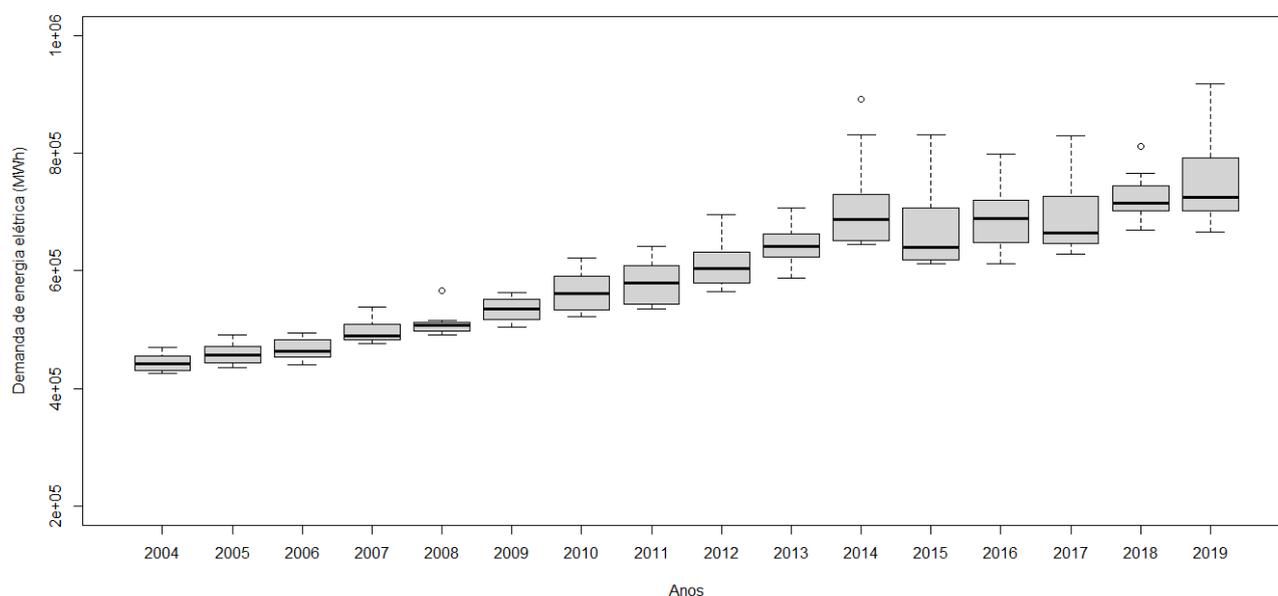


Figura 4.7 - Gráfico de caixa para cada ano da demanda mensal de energia elétrica da classe residencial/RS, 2004 - 19

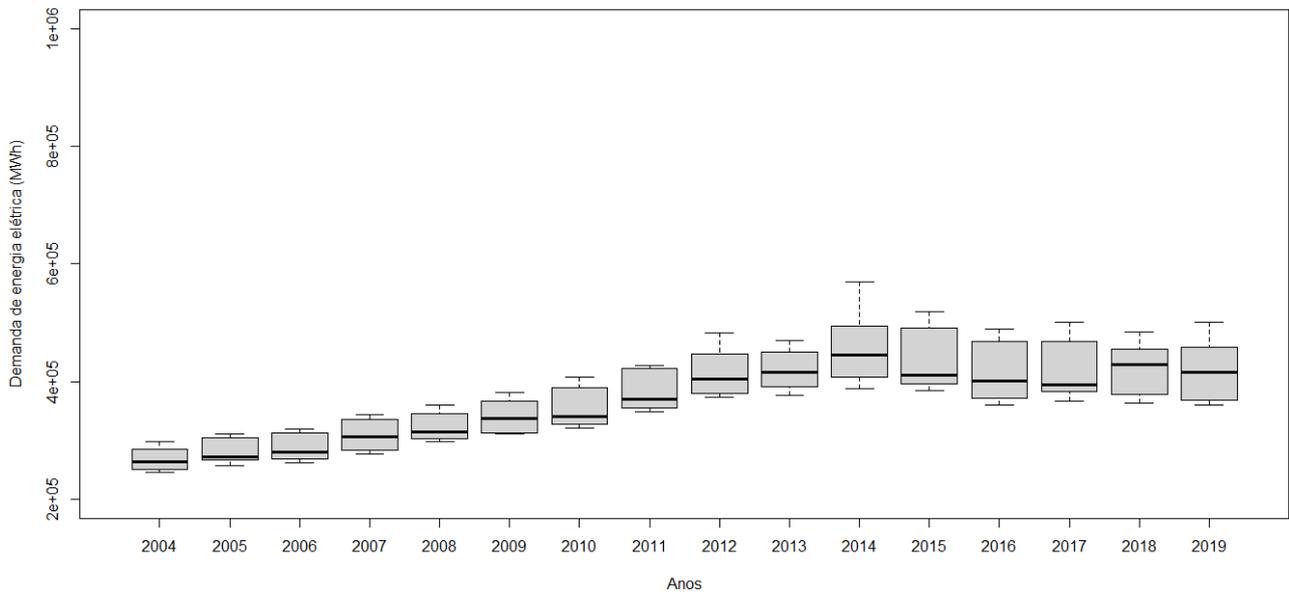


Figura 4.8 - Gráfico de caixa para cada ano da demanda mensal de energia elétrica da classe comercial/RS, 2004 - 19

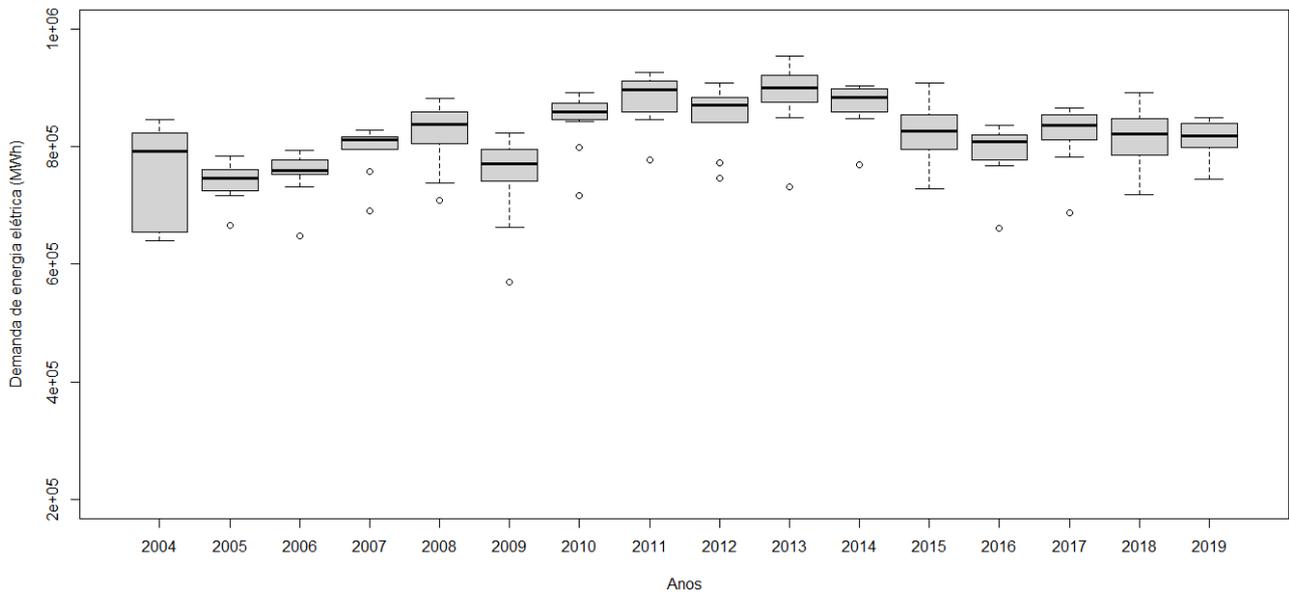


Figura 4.9 - Gráfico de caixa para cada ano da demanda mensal de energia elétrica da classe industrial/RS, 2004 - 19

Nestes gráficos nota-se a presença de possíveis valores discrepantes superiores na classe residencial nos anos de 2008, 2014 e 2018. Além da classe residencial, a classe industrial também apresentou vários possíveis valores fora do padrão inferiores nos anos de 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2016 e 2017, sendo importante ressaltar que os anos de 2007, 2010 e 2012 apresentam mais de um possível valor discrepante.

As figuras 4.10, 4.11 e 4.12 mostram os gráficos de caixa em cada mês para as classes residencial, comercial e industrial, respectivamente. É possível perceber que na classe residencial os dados encontram-se mais dispersos; na classe industrial, no mês de novembro, aparece um possível valor discrepante inferior, devendo sua influência ser analisada com cuidado para não prejudicar a modelagem da série temporal.

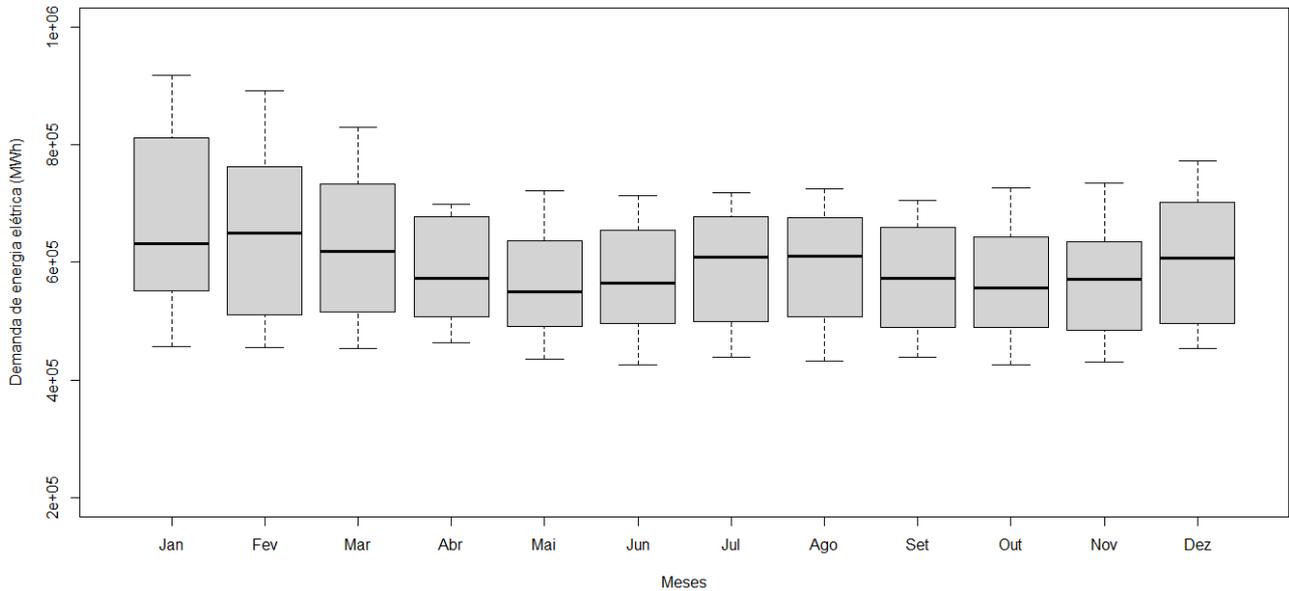


Figura 4.10 - Gráfico de caixa para cada mês ao longo dos anos da demanda mensal de energia elétrica da classe residencial/RS, 2004 – 19

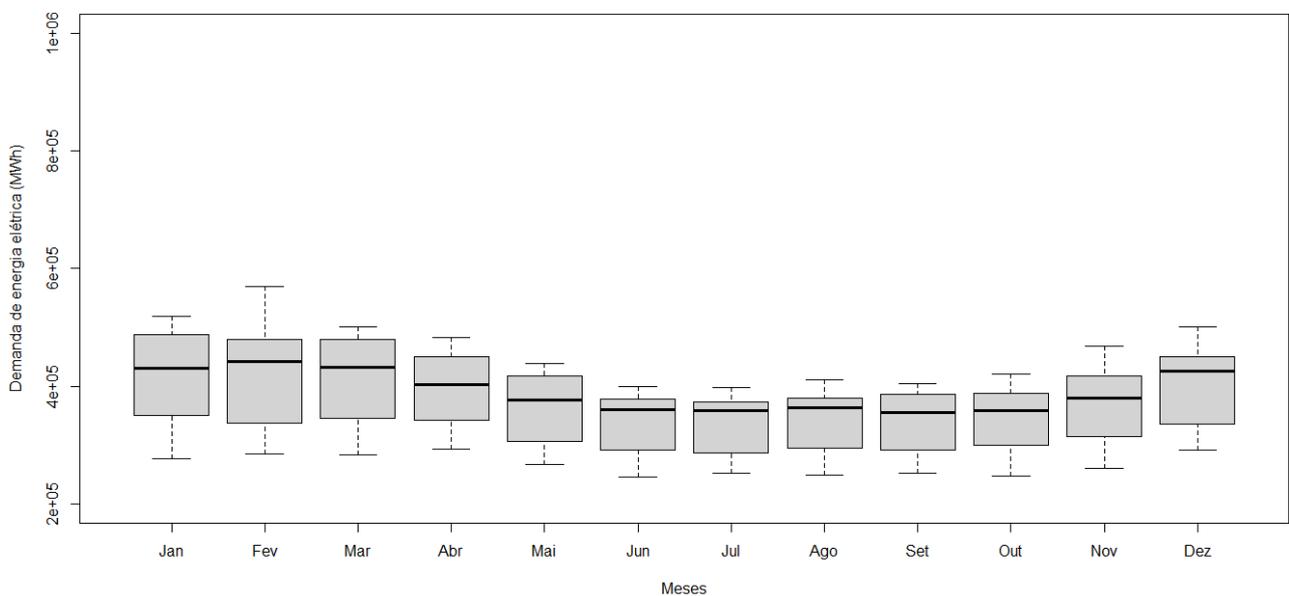


Figura 4.11 - Gráfico de caixa para cada mês ao longo dos anos da demanda mensal de energia elétrica da classe comercial/RS, 2004 - 19

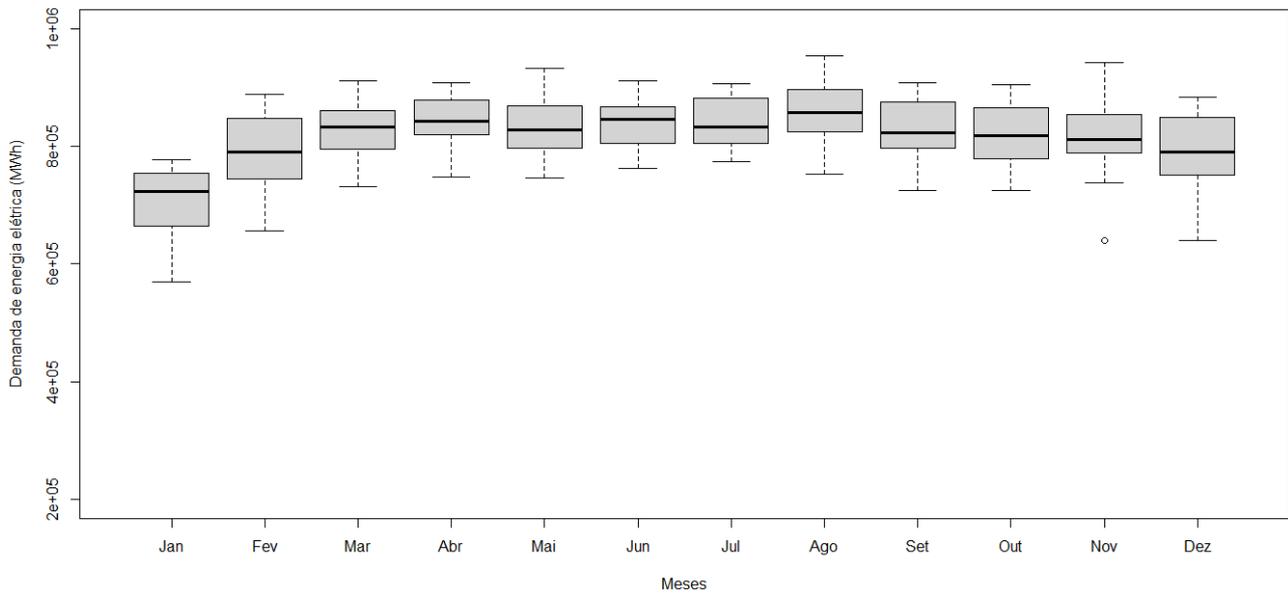


Figura 4.12- Gráfico de caixa para cada mês ao longo dos anos da demanda mensal de energia elétrica da classe industrial/RS, 2004 - 19

Com as análises gráficas apresentadas verifica-se que as séries estudadas têm um comportamento sazonal, como mostrado nos gráficos em linha, gráfico das subséries e boxplot. Entretanto por se tratar de uma avaliação gráfica, são aplicados testes para confirmar estas suposições.

A Tab. 4.2 apresenta o resultado do teste ANOVA com blocos para as três classes consumidoras, que testou a hipótese nula (H_0) de que não existe variação da média entre as observações do mesmo tratamento (meses) e entre os blocos (anos). Por meio do p-valor, utilizando nível de 0,05 de significância, rejeita-se H_0 , encontrando evidências de uma diferença significativa entre os tratamentos (meses), bem como entre os blocos (anos).

Tabela 4.2 - Resultado da ANOVA em blocos da demanda mensal de energia elétrica para as três classes consumidoras/RS, 2004 - 19

Classe	Medidas	Fonte de variação		
		Blocos	Tratamentos	Erros
Residencial	SQ	$1,94 \times 10^{12}$	$2,17 \times 10^{11}$	$1,60 \times 10^{11}$
	QM	$1,30 \times 10^{11}$	$1,97 \times 10^{10}$	$9,69 \times 10^8$
	gl	15	11	165
	f	133,80	20,38	
	pvalor	<0,0001*	<0,0001*	
Comercial	SQ	$6,89 \times 10^{11}$	$2,08 \times 10^{11}$	$4,87 \times 10^{10}$
	QM	$4,59 \times 10^{10}$	$1,89 \times 10^{10}$	$2,95 \times 10^8$
	gl	15	11	165
	f	155,67	64,31	
	pvalor	<0,0001*	<0,0001*	
Industrial	SQ	$4,00 \times 10^{11}$	$2,75 \times 10^{11}$	$1,70 \times 10^{11}$
	QM	$2,66 \times 10^{10}$	$2,50 \times 10^{10}$	$1,03 \times 10^9$
	gl	15	11	165
	f	25,86	24,27	
	pvalor	<0,0001*	<0,0001*	

Obs: * valor significativo para nível de significância de 5,0%.

Fonte: Elaborada pela autora.

A Tab. 4.3 apresenta o resultado da análise de resíduos do teste ANOVA com blocos. Nesta análise rejeitou-se H_0 para os testes de independência e normalidade, ou seja, os resíduos apresentam evidências de estarem autocorrelacionados e de não apresentarem distribuição normal. Entretanto, no teste de homocedasticidade, para as três classes consumidoras, não se rejeita a H_0 , não existindo evidência de resíduos heterocedásticos.

Desta forma, de acordo com a análise dos resíduos não foram satisfeitas as suposições do modelo teórico.

Tabela 4.3- Análise dos resíduos do teste ANOVA com blocos da demanda mensal de energia elétrica/RS, 2004 -19

Classe	Ljung-Box (p-valor)	ARCH (p-valor)	Jarque-Bera (p-valor)
Residencial	<0,0001*	0,1723	<0,0001*
Comercial	<0,0001*	0,0429*	0,0024*
Industrial	<0,0001*	0,8440	<0,0001*

Obs: * valor significativo para nível de significância de 5,0%.

Fonte: Elaborada pela autora.

Em decorrência dos resultados apresentados pelo teste ANOVA em blocos, realizou-se a aplicação do teste não paramétrico de Friedman, cujos resultados estão apresentados na Tab. 4.4.

Tabela 4.4 - Resultados do teste de Friedman da demanda mensal de energia elétrica para as três classes consumidoras/RS, 2004 - 19

Classe	Q	gl	p-valor
Residencial	117,62	11	<0,0001*
Comercial	152,11	11	<0,0001*
Industrial	87,94	11	<0,0001*

Obs: * valor significativo para nível de significância de 5,0%.

Fonte: Elaborada pela autora.

Nas três classes consumidoras rejeita-se a hipótese nula, ou seja, revelam-se evidências de que em pelo menos um dos meses a distribuição da demanda é diferente, o que reforça a ideia de presença de sazonalidade em todas as séries analisadas.

Seguindo a quase totalidade das modelagens da demanda de energia elétrica por métodos estatísticos, foram aplicadas transformações logarítmicas a fim de amenizar sua variabilidade. As figuras 4.13, 4.14, 4.15 mostram os gráficos do logaritmo da demanda mensal de energia elétrica para as classes residencial, comercial e industrial, respectivamente, em que se nota a diminuição da variabilidade, conforme esperado.

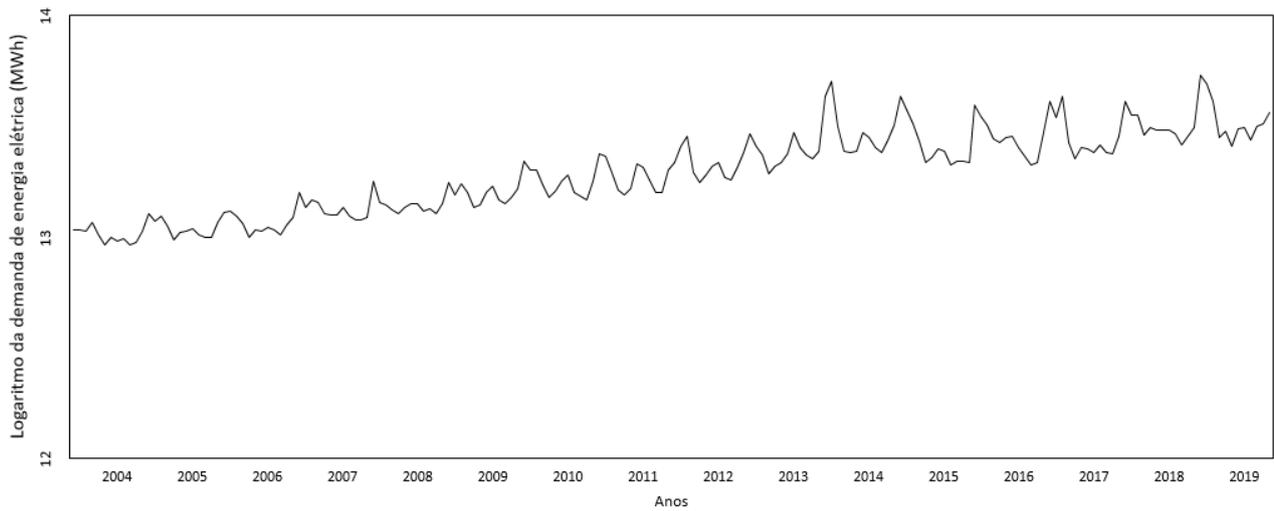


Figura 4.13 - Logaritmo da demanda mensal de energia elétrica da classe residencial/ RS, 2004 – 19

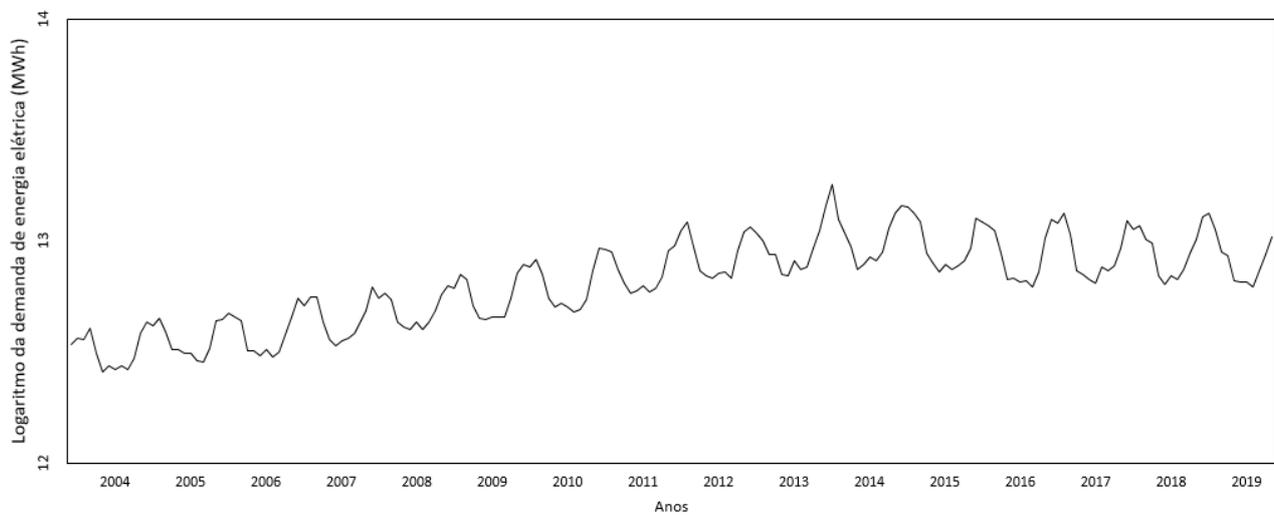


Figura 4.14 - Logaritmo da demanda mensal de energia elétrica da classe comercial/RS, 2004 – 19

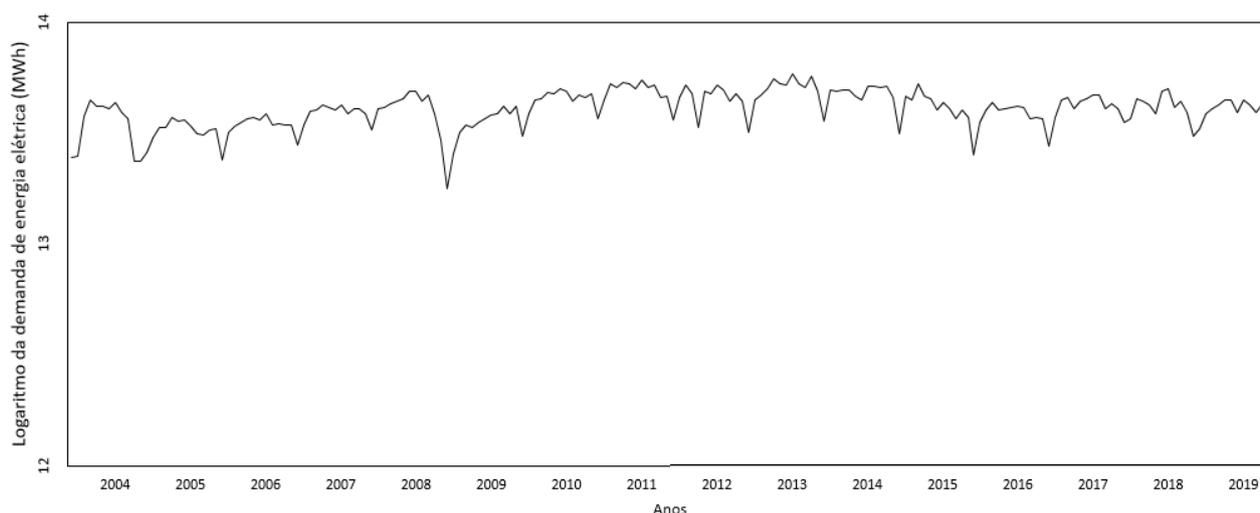


Figura 4.15 - Logaritmo da demanda mensal de energia elétrica da classe industrial/RS, 2004 – 19

A Tab. 4.5 apresenta o resultado do teste ADF aplicado aos dados de ajuste com a transformação logarítmica. Os valores apresentados são utilizados para avaliação da significância dos coeficientes δ , β_1 e β_2 , de acordo com as características de cada série e indicam a ordem de integração em que existem evidências de que a série se torna estacionária. Nota-se que na parte não sazonal, para todas as três classes, a série com os dados da demanda é estacionária na primeira diferença para modelo constante e tendência. Já a parte sazonal é estacionária em nível para modelo apenas com constante nas classes residencial e industrial. A parte sazonal dos dados da classe comercial é estacionária em primeira diferença, para modelo com constante e tendência.

Tabela 4.5 - Resultado do teste de estacionariedade da demanda mensal de energia elétrica para as três classes consumidoras/RS, 2004 - 18

Classe		Parte não sazonal			Parte sazonal			(d, D)
		$\delta (\tau)$	$\beta_1(\varphi_1, \varphi_2)$	$\beta_2(\varphi_3)$	$\delta (\tau)$	$\beta_1(\varphi_1, \varphi_2)$	$\beta_2(\varphi_3)$	
Dem _{res}	Crít.	-3,43	4,75	6,49	-2,88	4,63		(1, 0)
	Calc.	-3,91	5,11	7,66	-3,36	5,65		
Dem _{con}	Crít.	-3,43	4,75	6,49	-3,43	4,75	6,49	(1,1)
	Calc.	-14,12	66,47	99,71	-4,5	6,83	10,25	
Dem _{ind}	Crít.	-3,43	4,75	6,49	-2,88	4,63		(1,0)
	Calc.	-4,25	6,10	9,13	-3,48	6,12		

Fonte: Elaborada pela autora.

As figuras 4.16, 4.17 e 4.18 apresentam a FAC e FACP, para parte não sazonal e sazonal, das classes residencial, comercial e industrial, respectivamente. Nesta etapa investigou-se a ordem da componente de média móvel (MA) e da componente autoregressiva (AR) através dos correlogramas da FAC e FACP, construídos levando em consideração as respectivas ordens de integração sazonal e não sazonal.

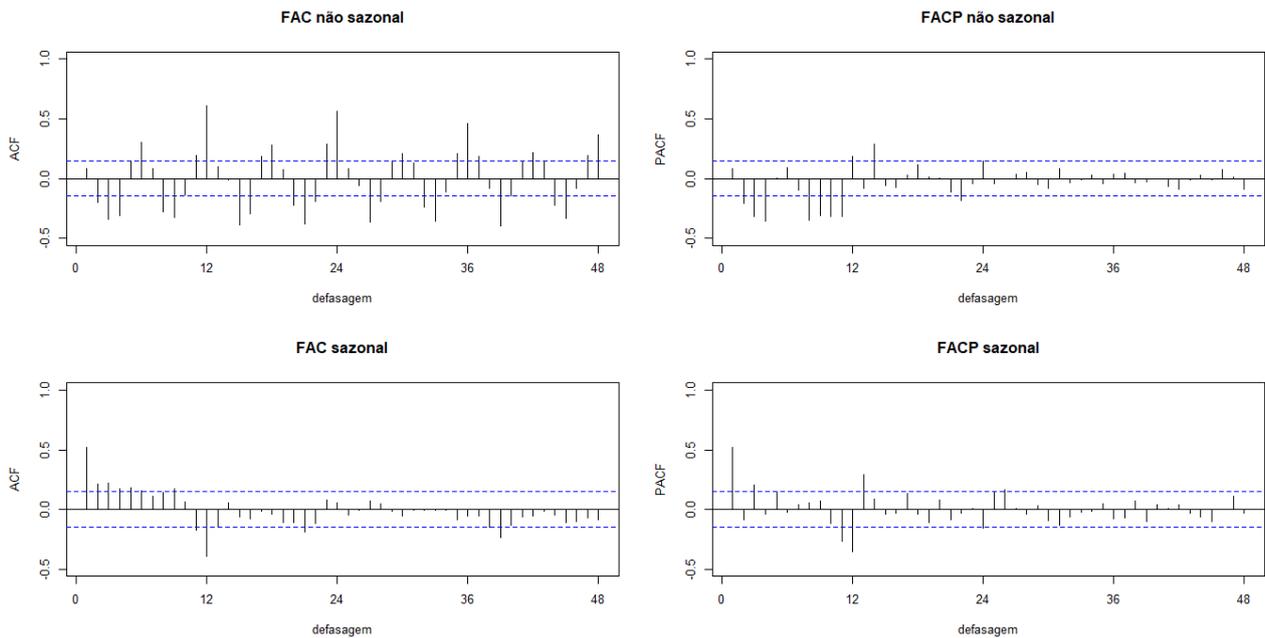


Figura 4.16 - FAC e FACP, da parte não sazonal e sazonal, do logaritmo da demanda mensal de energia elétrica da classe residencial/RS, 2004 - 18

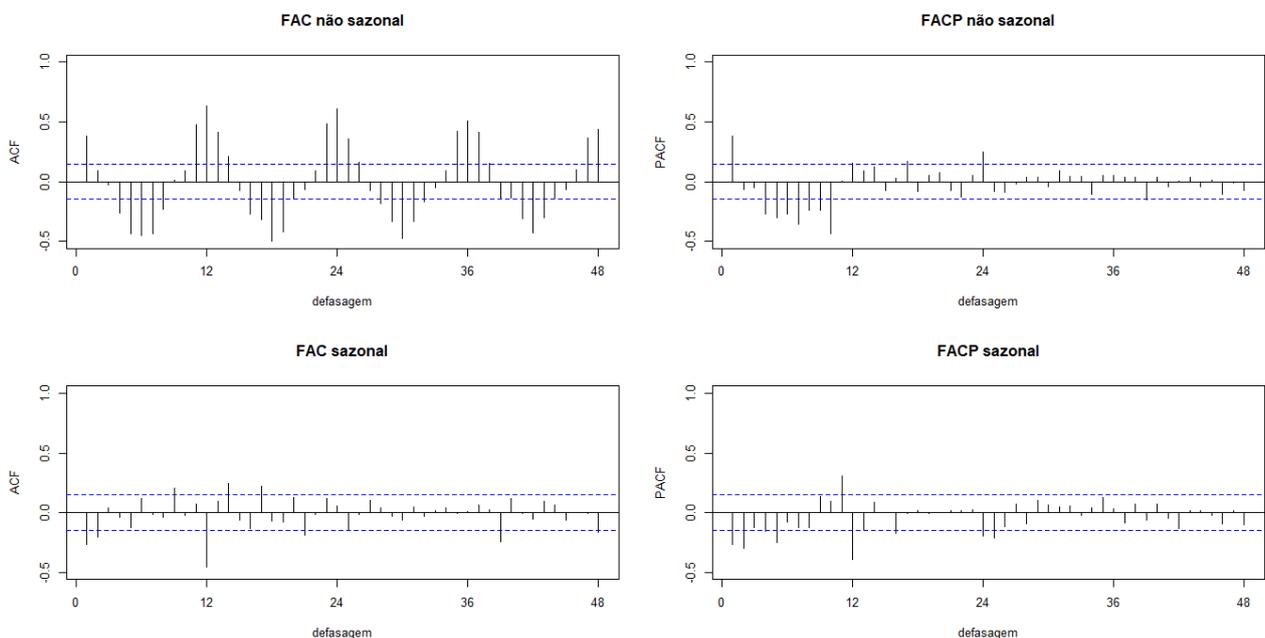


Figura 4.17 - FAC e FACP, da parte não sazonal e sazonal, do logaritmo da demanda mensal de energia elétrica da classe comercial/RS, 2004 - 18

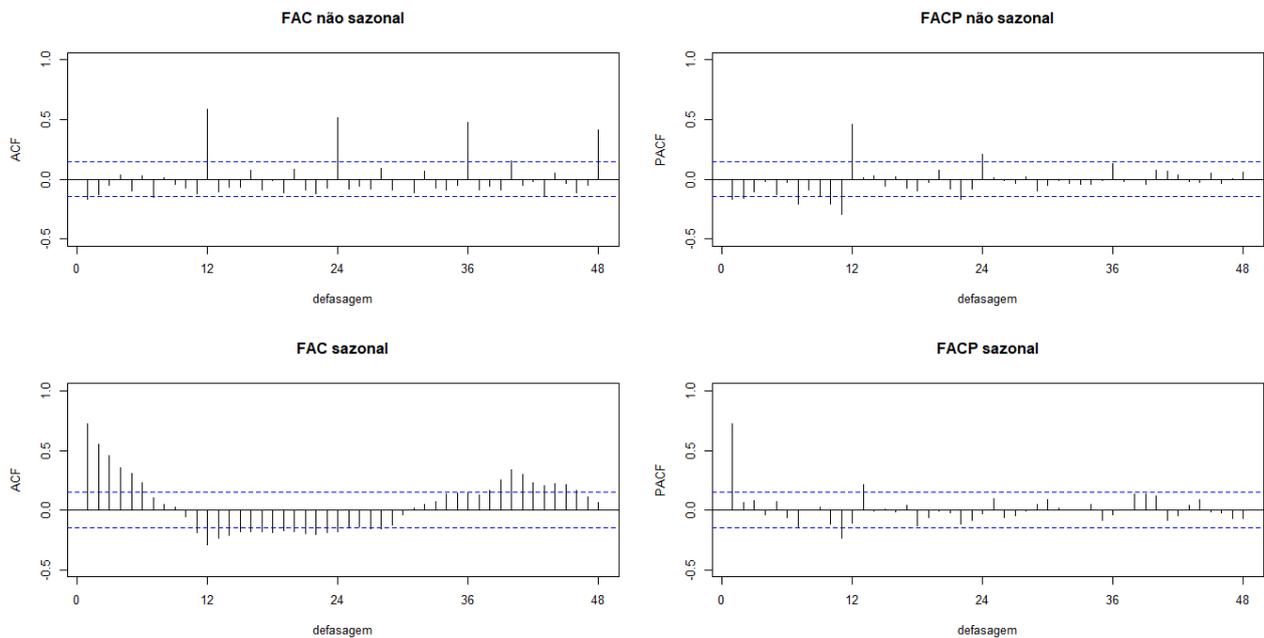


Figura 4.18 - FAC e FACP, da parte não sazonal e sazonal, do logaritmo da demanda mensal de energia elétrica da classe industrial/RS, 2004 - 18

Como a interpretação dos correlogramas não foi conclusiva, optou-se por estimar os coeficientes dos 256 modelos candidatos. Conforme apresentado na metodologia, foi avaliada a sua parcimônia através do BIC, sendo selecionados 13 modelos. As tabelas 4.6, 4.7, e 4.8 mostram esses modelos selecionados com respectivo BIC, bem como o resultado da averiguação da presença de autocorrelação, normalidade e homocedasticidade nos resíduos, o que foi feito para as classes residencial, comercial e industrial, respectivamente.

A Tabela 4.6 mostra que, entre os modelos selecionados para a classe residencial, os resíduos dos modelos SARIMA $(0,1,2)(1,0,1)_{12}$, SARIMA $(1,1,2)(1,0,1)_{12}$, SARIMA $(0,1,3)(1,0,1)_{12}$, SARIMA $(0,1,2)(1,0,2)_{12}$, SARIMA $(0,1,2)(2,0,1)_{12}$, SARIMA $(0,1,2)(2,0,0)_{12}$, SARIMA $(0,1,2)(1,0,3)_{12}$ e SARIMA $(2,1,2)(1,0,1)_{12}$ não apresentam evidências de autocorrelação, uma vez que a hipótese nula não pode ser rejeitada, e nem heterocedasticidade, pois a hipótese nula para este teste também não pode ser rejeitada. Com relação à normalidade, rejeitou-se a hipótese nula encontrando-se evidências de que os resíduos não provêm de uma população com distribuição normal.

Tabela 4.6 - Resultado do critério de informação e análise dos resíduos da demanda mensal de energia elétrica para classe residencial/RS, 2004 -18

Modelo SARIMA(p,d,q)(P,D,Q)_s	BIC	Ljung-Box (p-valor)	Jarque-Bera (p-valor)	ARCH (p-valor)
SARIMA (0,1,2)(1,0,1) ₁₂	-607,76	0,0207	<0,0001*	0,1035
SARIMA (1,1,1)(1,0,1) ₁₂	-603,77	0,0015*	<0,0001*	0,0215
SARIMA (1,1,2)(1,0,1) ₁₂	-602,84	0,0460	<0,0001*	0,1152
SARIMA (0,1,3)(1,0,1) ₁₂	-602,70	0,0316	<0,0001*	0,1109
SARIMA (0,1,2)(1,0,2) ₁₂	-602,57	0,0219	<0,0001*	0,0999
SARIMA (0,1,2)(2,0,1) ₁₂	-602,57	0,0223	<0,0001*	0,0982
SARIMA (0,1,2)(2,0,0) ₁₂	-600,65	0,0296	<0,0001*	0,4291
SARIMA (2,1,1)(1,0,1) ₁₂	-600,46	0,0027*	<0,0001*	0,0359
SARIMA (0,1,2)(3,0,0) ₁₂	-599,61	0,0058*	<0,0001*	0,4808
SARIMA (0,1,2)(1,0,3) ₁₂	-599,46	0,0270	<0,0001*	0,1419
SARIMA (2,1,2)(1,0,1) ₁₂	-599,07	0,0412	<0,0001*	0,0927
SARIMA (1,1,1)(2,0,1) ₁₂	-598,74	0,0010*	<0,0001*	0,0348
SARIMA (1,1,1)(1,0,2) ₁₂	-598,69	0,0011*	<0,0001*	0,0303

Obs: * valor significativo para nível de significância de 1,0%.

Fonte: Elaborada pela autora.

A Tab. 4.7 mostra que, entre os candidatos selecionados para a classe comercial, os modelos SARIMA (2,1,3)(0,1,1)₁₂, SARIMA (3,1,3)(0,1,1)₁₂, SARIMA (2,1,3)(0,1,2)₁₂ e SARIMA (1,1,2)(0,1,1)₁₂, satisfazem as suposições em relação às propriedades de autocorrelação, normalidade e homocedasticidade.

Tabela 4.7- Resultado do critério de informação e análise dos resíduos da demanda mensal de energia elétrica para classe comercial/RS, 2004 - 18

Modelo SARIMA (p,d,q)(P,D,Q)_s	BIC	Ljung-Box (p-valor)	Jarque-Bera (p-valor)	ARCH (p-valor)
SARIMA (0,1,2)(0,1,1) ₁₂	-615,90	0,0024*	0,0158	0,1780
SARIMA (2,1,3)(0,1,1) ₁₂	-615,54	0,4247	0,5800	0,3757
SARIMA (1,1,1)(0,1,1) ₁₂	-615,11	0,0010*	0,0107	0,1371
SARIMA (0,1,1)(0,1,1) ₁₂	-614,95	0,0010*	0,0137	0,2539
SARIMA (3,1,3)(0,1,1) ₁₂	-612,52	0,6662	0,4511	0,3580
SARIMA (2,1,3)(0,1,2) ₁₂	-611,55	0,5025	0,6298	0,4642
SARIMA (1,1,2)(0,1,1) ₁₂	-611,40	0,0176	0,0182	0,1996
SARIMA (0,1,3)(0,1,1) ₁₂	-611,05	0,0066*	0,0185	0,2039
SARIMA (0,1,2)(0,1,2) ₁₂	-610,78	0,0025*	0,0154	0,1744
SARIMA (0,1,2)(1,1,1) ₁₂	-610,78	0,0025*	0,0153	0,1742
SARIMA (2,1,1)(0,1,1) ₁₂	-610,26	0,0012*	0,0124	0,1691
SARIMA (0,1,1)(1,1,1) ₁₂	-610,09	0,0005*	0,0318	0,3027
SARIMA (0,1,1)(0,1,2) ₁₂	-610,04	0,0006*	0,0273	0,2914

Obs: * valor significativo para nível de significância de 1,0%.

Fonte: Elaborada pela autora.

A Tab. 4.8 mostra que todos os resíduos dos modelos selecionados para representar a demanda mensal de energia elétrica da classe industrial não apresentam evidências de autocorrelação e nem de heterocedasticidade. Entretanto, todos os modelos candidatos apresentam evidências de não normalidade, assim como ocorreu entre os modelos candidatos da classe residencial.

Tabela 4.8 - Resultado do critério de informação e análise dos resíduos da demanda mensal de energia elétrica para classe industrial/RS, 2004 - 18

Modelo SARIMA (p,d,q)(P,D,Q)_s	BIC	Ljung-Box (p-valor)	Jarque-Bera (p-valor)	ARCH (p-valor)
SARIMA (1,1,1)(1,0,1) ₁₂	-572,94	0,9917	<0,0001*	0,9115
SARIMA (0,1,2)(1,0,1) ₁₂	-570,53	0,9980	<0,0001*	0,9650
SARIMA (0,1,1)(1,0,1) ₁₂	-569,91	0,9924	<0,0001*	0,9645
SARIMA (1,1,2)(1,0,1) ₁₂	-568,88	0,9985	<0,0001*	0,9335
SARIMA (2,1,1)(1,0,1) ₁₂	-568,54	0,9971	<0,0001*	0,9211
SARIMA (1,1,1)(1,0,2) ₁₂	-568,05	0,9937	<0,0001*	0,9137
SARIMA (1,1,1)(2,0,1) ₁₂	-568,00	0,9933	<0,0001*	0,9144
SARIMA (2,1,0)(1,0,1) ₁₂	-566,80	0,9995	<0,0001*	0,9171
SARIMA (1,1,3)(1,0,1) ₁₂	-566,23	0,9998	<0,0001*	0,9513
SARIMA (2,1,2)(1,0,1) ₁₂	-566,02	0,9985	<0,0001*	0,9523
SARIMA (0,1,3)(1,0,1) ₁₂	-565,63	0,9958	<0,0001*	0,9627
SARIMA (3,1,1)(1,0,1) ₁₂	-565,63	0,9999	<0,0001*	0,9450
SARIMA (0,1,2)(1,0,2) ₁₂	-565,43	0,9984	<0,0001*	0,9652

Obs: * valor significativo para nível de significância de 1,0%.

Fonte: Elaborada pela autora.

Dos modelos apresentados e avaliados nas tabelas 4.6, 4.7 e 4.8, foram selecionados aqueles que não apresentam resíduos autocorrelacionados para analisar sua capacidade preditiva, sendo as violações da normalidade relaxadas (BUENO, 2011). As tabelas 4.9, 4.10 e 4.11 apresenta as medidas de acurácia: raiz do erro quadrático médio (RMSE), o erro absoluto médio (MAE), o erro absoluto percentual médio (MAPE) e o U de Theil dos modelos selecionados.

De acordo com a Tab. 4.9, o modelo SARIMA (1,1,2)(1,0,1)₁₂ apresenta os menores valores de RMSE e MAE e o modelo SARIMA (0,1,2)(1,0,3)₁₂ os menores valores para MAPE e U de Theil.

Dessa maneira, tem-se dois modelos candidatos que podem ser escolhidos para prever a demanda de energia elétrica da classe residencial.

Tabela 4.9 - Indicadores de qualidade para validação dos modelos da classe residencial/RS, 2019

Modelo SARIMA (p,d,q)(P,D,Q)_s	RMSE	MAE	MAPE(%)	U de Theil
SARIMA (0,1,2)(1,0,1) ₁₂	42705,9	33858,63	4,3497	0,6704
SARIMA (1,1,2)(1,0,1) ₁₂	42384,8*	33597,64*	4,3236	0,6660
SARIMA (0,1,3)(1,0,1) ₁₂	42540,9	33721,56	4,3361	0,6681
SARIMA (0,1,2)(1,0,2) ₁₂	42724,1	33905,36	4,3559	0,6702
SARIMA (0,1,2)(2,0,1) ₁₂	42731,1	33926,53	4,3587	0,6701
SARIMA (0,1,2)(2,0,0) ₁₂	49539,3	36280,75	4,6060	0,7335
SARIMA (0,1,2)(1,0,3) ₁₂	43689,9	33601,48	4,3071*	0,6572*
SARIMA (2,1,2)(1,0,1) ₁₂	43124,9	34213,63	4,3887	0,6735

Obs: * significa menor valor.

Fonte: Elaborada pela autora.

Os dados da Tab. 4.10 mostram que, assim como na classe residencial, houve discordância entre os indicadores da capacidade preditiva, sendo selecionado o modelo SARIMA (2,1,3)(0,1,2)₁₂, com menores valores de MAE e MAPE, e o SARIMA (1,1,2)(0,1,1)₁₂, com menores valores de RMSE e U de Theil.

Tabela 4.10 - Indicadores de qualidade para validação dos modelos da classe comercial, 2019

Modelo SARIMA (p,d,q)(P,D,Q)_s	RMSE	MAE	MAPE (%)	U de Theil
SARIMA (2,1,3)(0,1,1) ₁₂	17346,93	13680,55	3,4275	0,6851
SARIMA (3,1,3)(0,1,1) ₁₂	16634,79	13087,50	3,2665	0,6535
SARIMA (2,1,3)(0,1,2) ₁₂	16222,53	12220,88*	3,0611*	0,6346
SARIMA (1,1,2)(0,1,1) ₁₂	16069,45*	12654,81	3,1156	0,6210*

Obs: * significa menor valor.

Fonte: Elaborada pela autora.

Para a classe industrial também foram identificados dois modelos com menores valores entre os quatro indicadores da capacidade preditiva, conforme apresentado na Tab. 4.11. O modelo SARIMA (1,1,1)(1,0,2)₁₂, com menor valor de RMSE, e o SARIMA(1,1,1)(2,0,1)₁₂, com menor valor de MAE, MAPE e U de Theil.

Tabela 4.11 - Indicadores de qualidade para validação dos modelos da classe industrial/RS, 2019

Modelo SARIMA (p,d,q)(P,D,Q)_s	RMSE	MAE	MAPE(%)	U de Theil
SARIMA (1,1,1)(1,0,1) ₁₂	35772,45	28356,60	3,5214	0,8387
SARIMA (0,1,2)(1,0,1) ₁₂	43542,53	38462,82	4,7377	1,1265
SARIMA (0,1,1)(1,0,1) ₁₂	59011,64	56089,57	6,8914	1,6270
SARIMA (1,1,2)(1,0,1) ₁₂	36290,04	29103,47	3,6065	0,8784
SARIMA (2,1,1)(1,0,1) ₁₂	36454,65	29190,28	3,6196	0,8771
SARIMA (1,1,1)(1,0,2) ₁₂	34964,48*	28123,88	3,4968	0,8284
SARIMA (1,1,1)(2,0,1) ₁₂	34985,92	28042,61*	3,4865*	0,8263*
SARIMA (2,1,0)(1,0,1) ₁₂	59186,94	56338,22	6,9203	1,6294
SARIMA (1,1,3)(1,0,1) ₁₂	35788,94	28935,03	3,5786	0,8595
SARIMA (2,1,2)(1,0,1) ₁₂	35578,76	28494,18	3,5309	0,8411
SARIMA (0,1,3)(1,0,1) ₁₂	41780,77	36299,84	4,4768	1,0600
SARIMA (3,1,1)(1,0,1) ₁₂	36511,14	29502,95	3,6464	0,8961
SARIMA (0,1,2)(1,0,2) ₁₂	43619,09	38479,98	4,7394	1,1361

Obs: * significa menor valor.

Fonte: Elaborada pela autora.

Como os indicadores da capacidade preditiva dos modelos identificaram dois modelos nas três classes consumidoras, para a identificação do melhor foram considerados os seguintes critérios: valor observado da série dentro do intervalo de confiança, menor valor de RMSE (segundo Morettin e Toloi (2018), este indicador é o mais frequentemente utilizado) e menor valor de BIC.

O modelo SARIMA (1,1,2)(1,0,1)₁₂ apresentou apenas um valor observado fora do intervalo de confiança, enquanto que o modelo SARIMA (0,1,2)(1,0,3)₁₂ apresentou dois valores observados fora do intervalo, conforme mostrado na Tab. 4.12. Além disso, o modelo SARIMA (1,1,2)(1,0,1)₁₂

tem o menor valor de RMSE e BIC, sendo o escolhido para prever a demanda mensal de energia elétrica da classe consumidora residencial.

Tabela 4.12 - Modelo SARIMA (1,1,2)(1,0,1)₁₂, valores observados, previstos e intervalo de confiança da demanda mensal de energia elétrica para a classe residencial/RS, 2019

Meses/2019	Valor observado	Valor previsto	IC-95%	
Janeiro	917352*	829988,3	767465,4	897604,7
Fevereiro	881917	806667,5	735735,5	884437,9
Março	813091	801159,7	729701,8	879615,3
Abril	692676	723932,5	657906,0	796585,4
Maiο	712893	710890,0	644738,6	783828,8
Junho	666041	721231,8	652803,5	796833,0
Julho	718333	732002,6	661248,7	810327,3
Agosto	725017	726868,5	655343,2	806200,3
Setembro	683906	713207,4	641807,6	792550,2
Outubro	726320	692325,1	621855,9	770779,9
Novembro	735331	702865,9	630168,4	783949,8
Dezembro	772078	743183,8	665117,0	830413,6

Obs: *valor observado fora do IC-95.

Fonte: Elaborada pela autora.

O modelo SARIMA (2,1,3)(0,1,2)₁₂ apresenta um valor observado fora do intervalo de confiança, enquanto o modelo SARIMA (1,1,2)(0,1,1)₁₂ nenhum, conforme mostrado na Tab. 4.13, além de apresentar menor RMSE (Tab. 4.10). Ademais, embora o primeiro apresente menor BIC, os valores desta medida são muito próximos, razão pela qual é mais razoável escolher o modelo SARIMA (1,1,2)(0,1,1)₁₂ para prever a demanda de energia elétrica da classe consumidora comercial.

Tabela 4.13 - Modelo SARIMA (1,1,2)(0,1,1)₁₂, valores observados, previstos e intervalo de confiança da demanda mensal de energia elétrica para a classe comercial/RS, 2019

Meses/2019	Valor observado	Previsão	IC-95%	
Janeiro	493023	486446,2	453880,3	521348,7
Fevereiro	501161	484511,7	447990,0	524010,8
Março	465609	482721,9	445168,9	523442,7
Abril	421338	455169,7	418156,8	495458,7
Maiο	413988	416872,3	381722,7	455258,6
Junho	370281	384605,1	351028,8	421393,0
Julho	367417	379213,3	345029,7	416783,6
Agosto	367663	385437,9	349632,7	424909,8
Setembro	360334	386056,0	349167,2	426842,1
Outubro	387465	390109,8	351828,7	432556,1
Novembro	416506	414787,4	373048,0	461196,9
Dezembro	451233	452055,2	405467,4	503995,8

Fonte: Elaborada pela autora.

Ambos os modelos, SARIMA (1,1,1)(1,0,2)₁₂, e SARIMA(1,1,1)(2,0,1)₁₂, apresentam apenas um valor observado fora do respectivo intervalo de confiança, conforme mostrado na Tab. 4.14. Entretanto, o modelo SARIMA (1,1,1)(1,0,2)₁₂ apresenta os menores valores para o indicador RMSE e para o BIC, acarretando na sua escolha para prever a demanda mensal de energia elétrica da classe consumidora industrial.

Tabela 4.14 - Modelo SARIMA (1,1,1)(1,0,2)₁₂, valores observados, previstos e intervalo de confiança da demanda mensal de energia elétrica para a classe industrial/RS, 2019

Meses/2019	Valor observado	Valor previsto	IC-95%	
Janeiro	744672*	669062,8	613604,6	729533,5
Fevereiro	797519	747439,7	671600,4	831842,9
Março	812329	797839,5	709763,5	896845,0
Abril	828741	813094,9	719231,0	919208,6
Maiο	847838	802319,1	707129,1	910323,0
Junho	848682	804320,9	707089,3	914922,8
Julho	800584	824433,2	723361,2	939627,7
Agosto	848966	841819,7	737440,1	960973,4
Setembro	825402	814427,6	712461,7	930986,6
Outubro	801219	807062,3	705141,6	923714,6
Novembro	831736	799509,7	697736,3	916128,1
Dezembro	779373	767630,2	669180,3	880563,9

Obs: *valor observado fora do IC-95

Fonte: Elaborada pela autora.

As figuras 4.19, 4.20 e 4.21 mostram os gráficos que representam os modelos SARIMA (1,1,2)(1,0,1)₁₂, SARIMA (1,1,2)(0,1,1)₁₂ e SARIMA (1,1,1)(1,0,2)₁₂, identificados como aqueles com melhor capacidade preditiva para as classes residencial, comercial e industrial, respectivamente. É possível perceber que os valores observados encontram-se dentro dos intervalos de confiança quase que em sua totalidade, indicando que os modelos são satisfatórios.

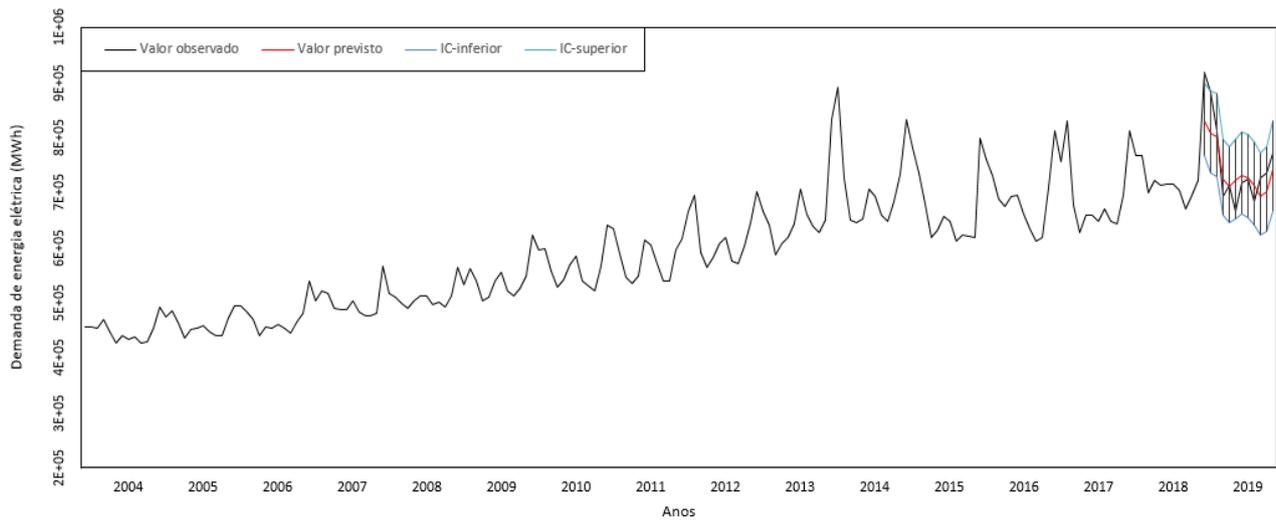


Figura 4.19 - Gráfico do modelo de previsão SARIMA (1,1,2) (1,0,1)12 para demanda mensal de energia elétrica da classe residencial/RS, 2004 - 19

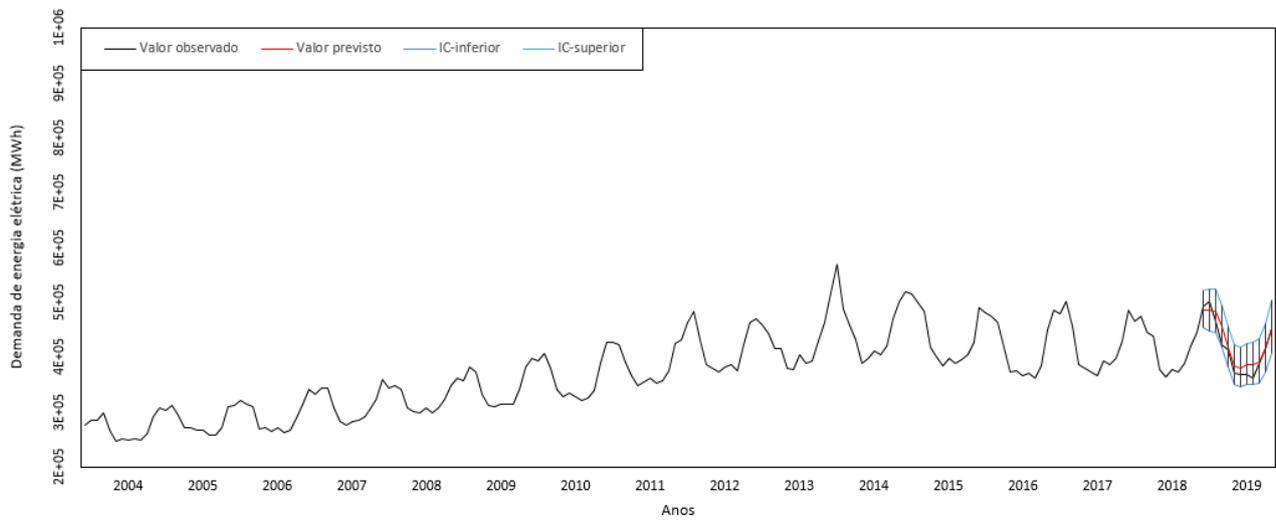


Figura 4.20 - Gráfico do modelo de previsão SARIMA (1,1,2) (0,1,1)12 para demanda mensal de energia elétrica da classe comercial/RS, 2004 - 19

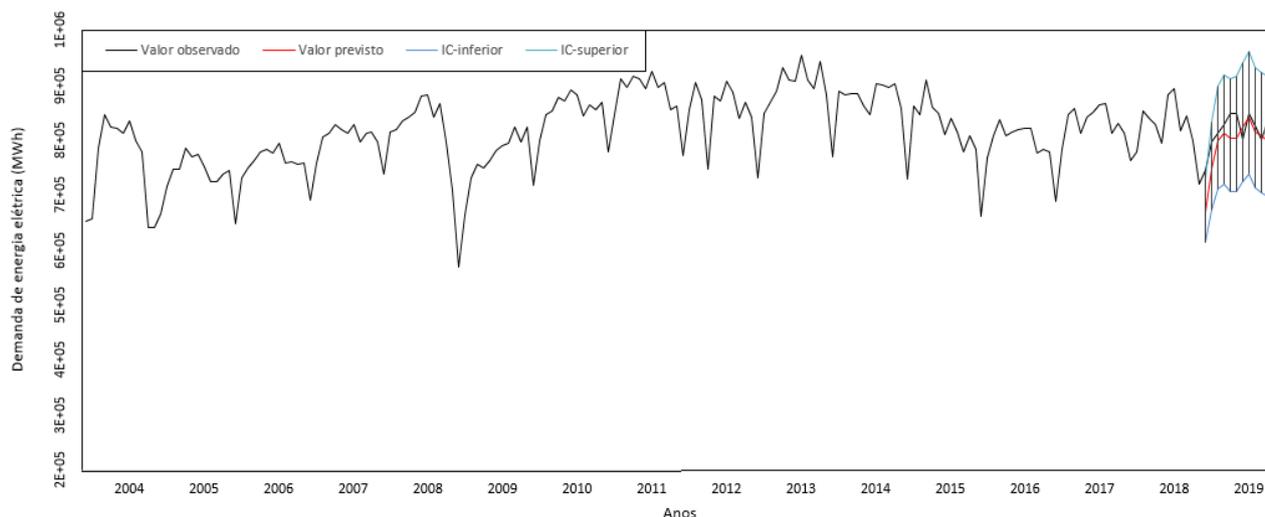


Figura 4.21 - Gráfico do modelo de previsão SARIMA (1,1,1) (1,0,2)12 para demanda mensal de energia elétrica da classe industrial/RS, 2004 - 19

Com todos os resultados apresentados é possível fazer uma análise comparativa entre os três modelos. As classes residencial e comercial apresentaram os mesmos valores de (p,d,q) na parte não sazonal (1,1,2) e diferentes valores nos parâmetros (P,D,Q) . Também pode-se perceber que nas três séries a parte não sazonal ficou estacionária na primeira diferente. Já na parte sazonal, apenas a classe comercial apresentou estacionariedade em primeira diferença, enquanto as classes residencial e comercial são estacionárias em nível.

Outra observação importante sobre os três modelos é que o modelo da classe comercial na parte sazonal depende apenas dos valores de erros passados ($Q=1$). Ademais, todos os outros parâmetros dos modelos selecionados dependem de valores passados (p e P) e de erros passados (q e Q) para fazer previsões.

Comparando os indicadores de qualidade entre os três modelos selecionados, nota-se que a classe comercial apresentou os menores valores, sendo também o modelo da classe comercial aquele que apresentou todos os valores previstos dentro do intervalo de confiança. Desta forma, pode-se supor que o modelo da classe comercial tem uma capacidade preditiva melhor se comparado aos modelos das classes residencial e industrial.

Para finalizar esta etapa são apresentados na Figura 4.22 os gráficos histograma e box plot dos resíduos para os modelos selecionados, pois rejeitou-se H_0 do teste de normalidade nas classes residencial e industrial, dando evidências de não normalidade. Pode-se perceber que nas três classes existem possíveis valores fora do padrão para os resíduos.

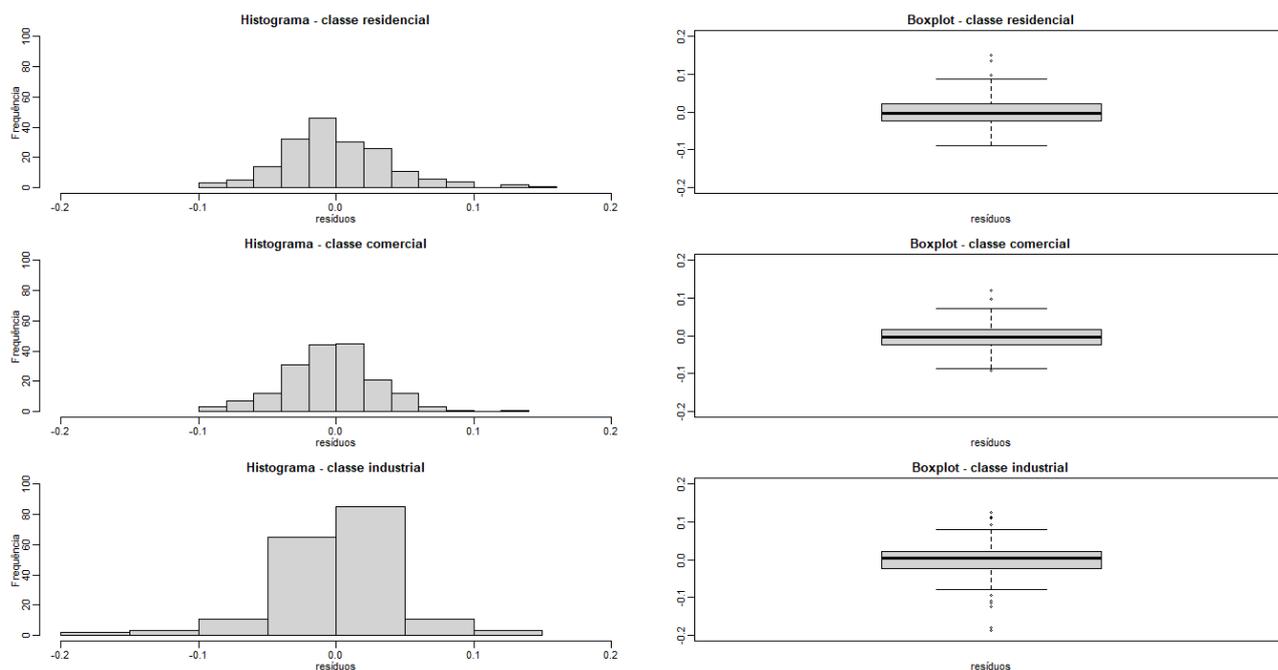


Figura 4.22 - Análise dos resíduos dos modelos selecionados para as três classes consumidoras/RS, 2004 - 18

Cabe ressaltar que futuros trabalhos podem ser desenvolvidos, usando técnicas mais consistentes, para verificar se estes valores são realmente discrepantes e, em caso positivo, ajustá-los, com objetivo de melhorar a modelagem destas séries e, conseqüentemente, a acurácia de suas previsões.

4.2. Refinamento do modelo: SARIMAX

As figuras 4.23, 4.24 e 4.25 mostram os gráficos em linhas e os gráficos das subséries sazonais do número de consumidores das classes residencial, comercial e industrial, respectivamente. Percebe-se que a classe residencial, Fig. 4.23(a), teve um crescimento do número de consumidores ao longo dos anos. Na classe comercial, Fig. 4.24(a), detecta-se um crescimento, entretanto de maneira mais irregular, diferente da residencial com crescimento gradativo. Com relação a classe industrial, Fig. 4.25(a), verifica-se um decaimento da série com perfil irregular ao longo dos anos.

Com relação ao gráfico das subséries sazonais nenhuma série apresentou meses com valores de média maiores ou menores, muito evidentes, em relação a outros meses. Nas classes residencial e comercial, Fig.4.23(b) e Fig. 4.24(b), as médias entre os meses mantiveram-se com valores próximos. Já na classe industrial, Fig. 4.25(b) tem-se um leve decaimento das médias entre os meses de agosto até dezembro.

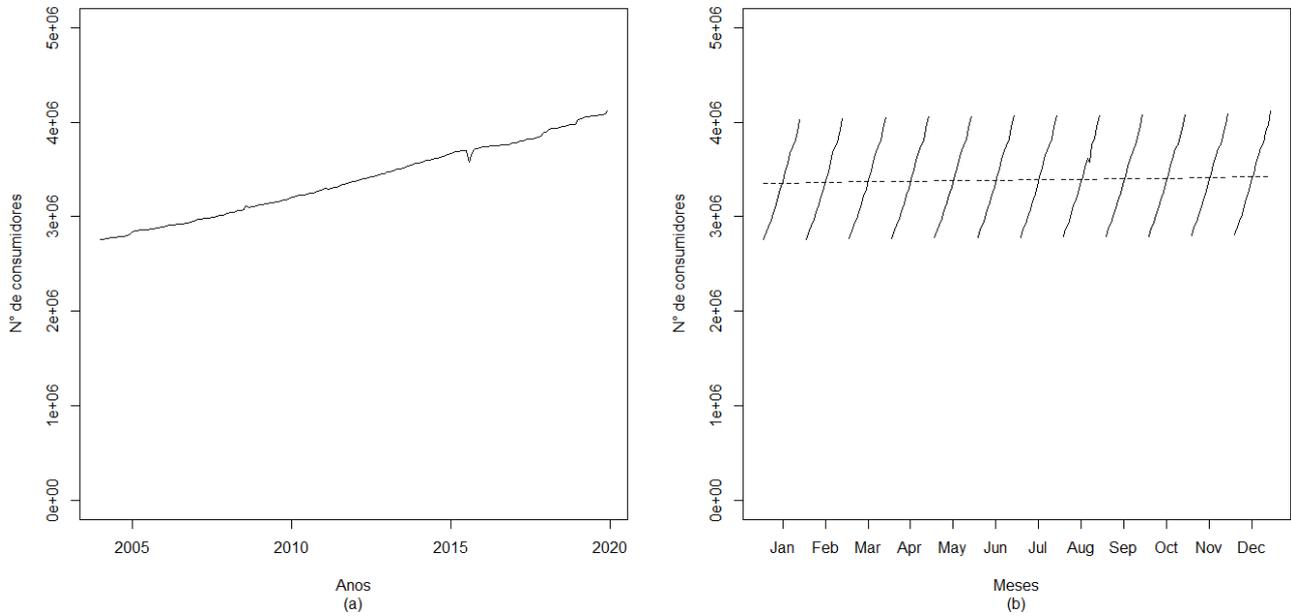


Figura 4.23 - Número de consumidores da classe residencial (a) em linha (b) subséries sazonais/RS, 2004 - 19

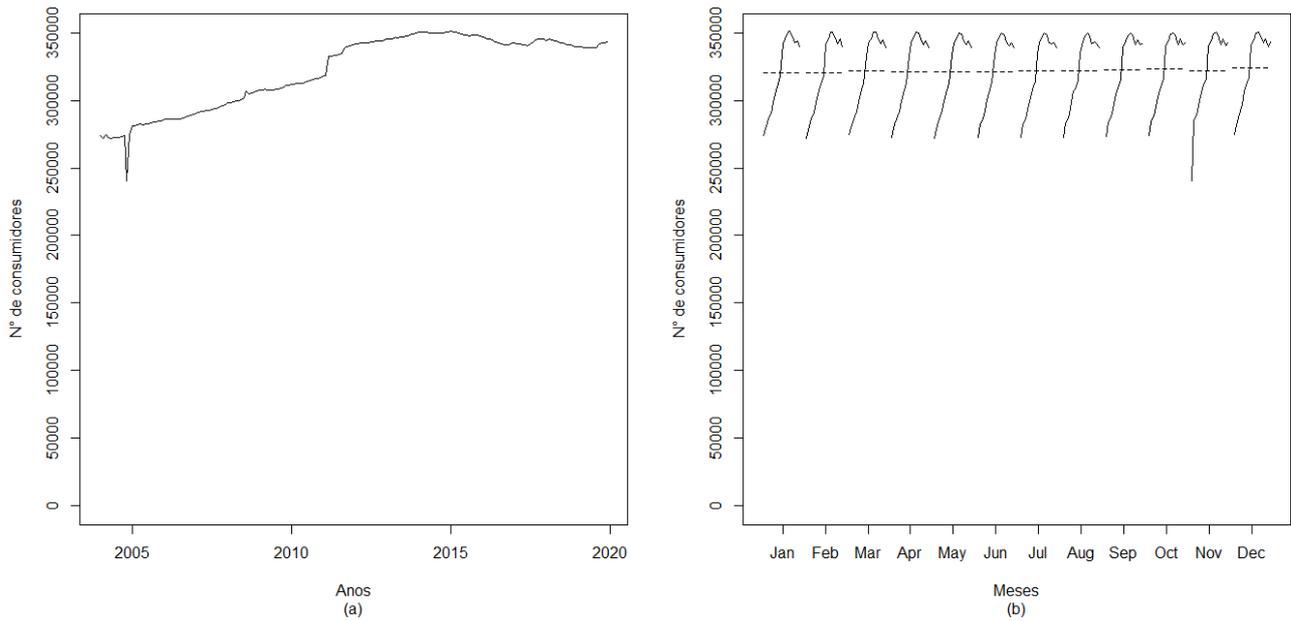


Figura 4.24 - Número de consumidores da classe comercial (a) em linha (b) subséries sazonais/RS, 2004 - 19

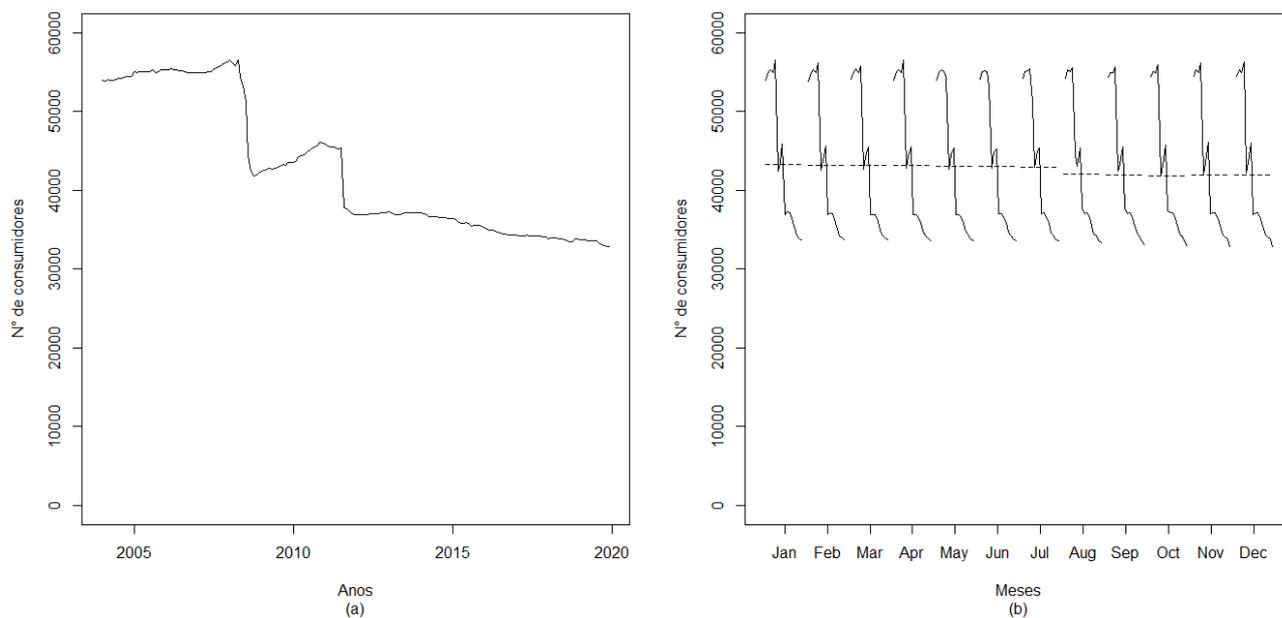


Figura 4.25 - Número de consumidores da classe industrial (a) em linha (b) subséries sazonais/RS, 2004 - 19

As figuras 4.26, 4.27 e 4.28 mostram os gráficos em linhas e os gráficos das subséries sazonais da tarifa de energia elétrica para as classes residencial, comercial e industrial, respectivamente. Nota-se, nas Fig. 4.26(a), Fig. 4.27(a) e Fig. 4.28(a), que todas as três séries, a partir de 2014, apresentaram um aumento expressivo nos valores das tarifas de energia elétrica.

O gráfico das subséries sazonais da classe residencial (Fig. 4.26(b)) mostra que as médias são bem próximas entre os meses, com leve aumento entre os meses de agosto até dezembro. Na classe comercial (Fig. 4.27 (b)) tem-se uma diferença maior entre as tarifas médias mensais, com aumento entre os meses de agosto até outubro. Na série da tarifa industrial, apresentada na Fig. 4.28 (b), é possível notar uma diferença maior entre as tarifas médias mensais, se comparada com as das outras classes.

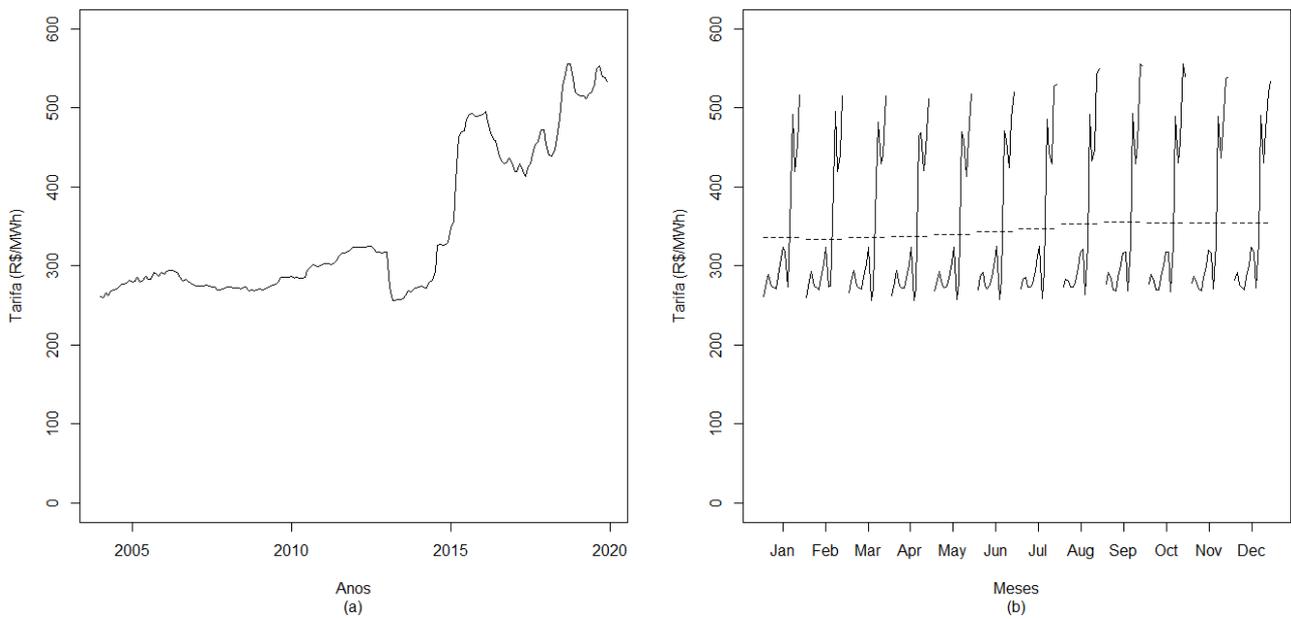


Figura 4.26 - Tarifa mensal de energia elétrica para a classe residencial (a) em linha (b) subséries sazonais/RS, 2004 - 19

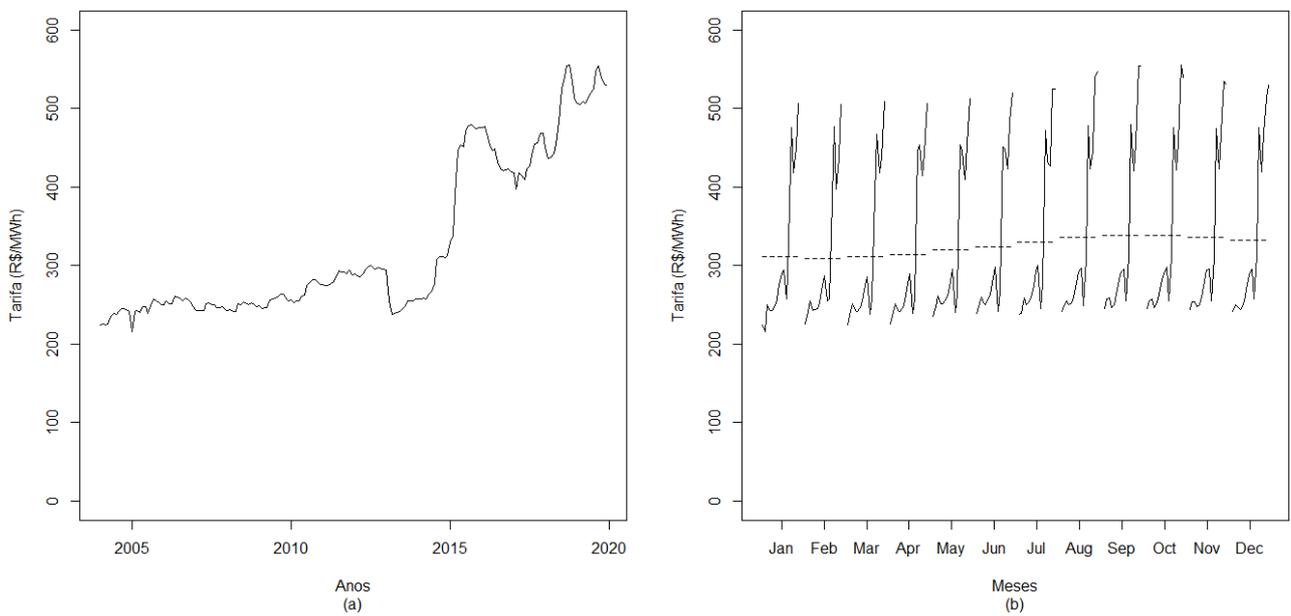


Figura 4.27 - Tarifa mensal de energia elétrica para a classe comercial (a) em linha (b) subséries sazonais/RS, 2004 - 19

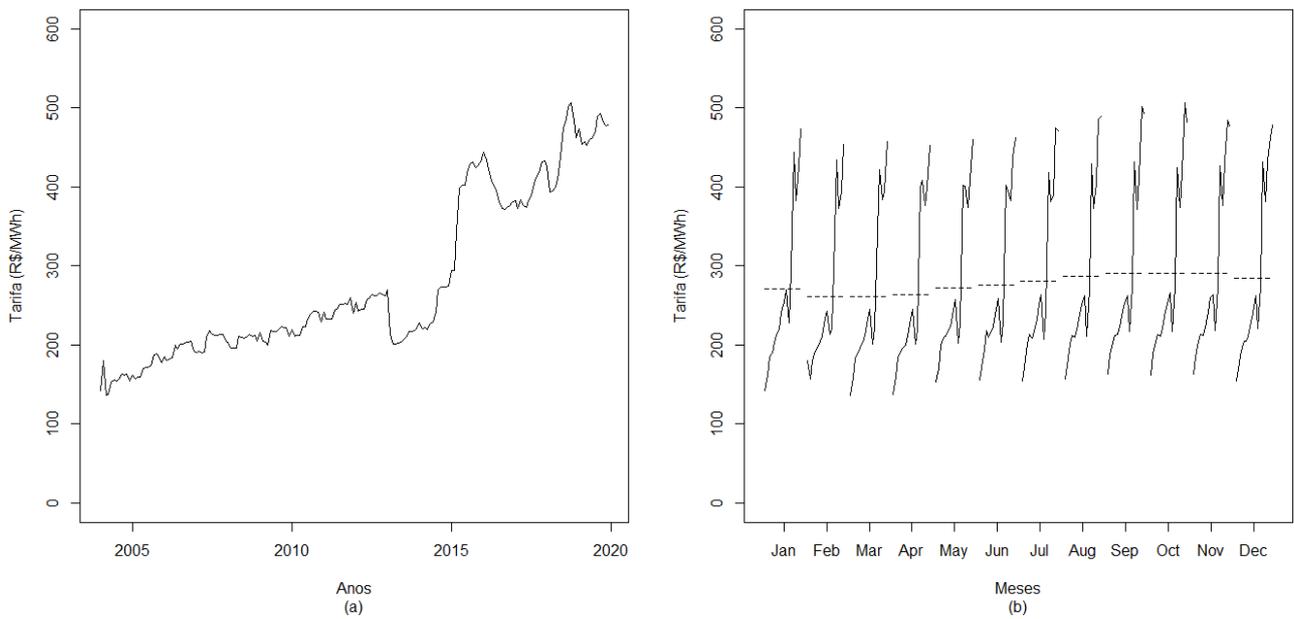


Figura 4.28 - Tarifa mensal de energia elétrica para a classe industrial (a) em linha (b) subséries sazonais/RS, 2004 - 19

A Fig. 4.29 mostra o gráfico em linhas e o gráfico das subséries sazonais da temperatura, uma variável comum às três classes consumidoras. Na Fig. 4.29(a) pode-se notar picos de sazonalidade, que são características dos meses de maior temperatura. Na Fig. 4.29(b) pode-se perceber que as médias maiores desta série são nos meses de janeiro, fevereiro e dezembro, sendo estes os meses que contemplam a estação de verão no estado.

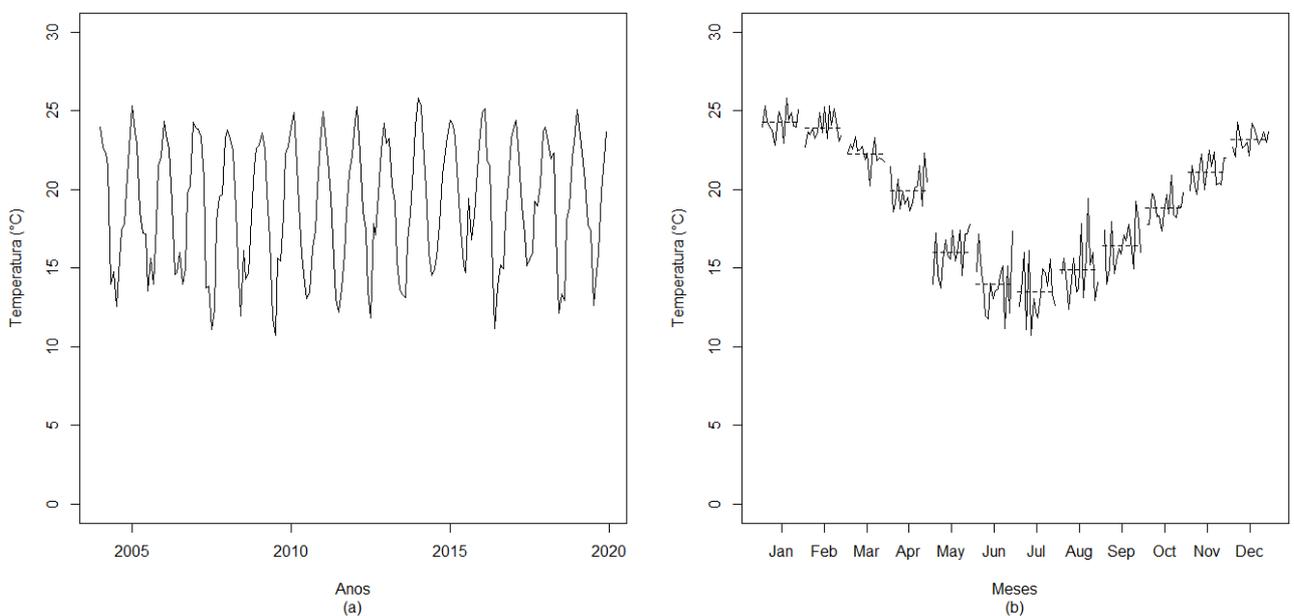


Figura 4.29 - Gráfico mensal da temperatura (a) em linha (b) subséries sazonais/RS, 2004 - 19

A Tab. 4.15 mostra as medidas resumo do número de consumidores, da tarifa de energia elétrica e da temperatura, para as três classes consumidoras.

Com relação ao número de consumidores a classe que apresenta maior resultado é a residencial e menor, a classe industrial. A maior variação relativa ocorreu na classe industrial e a menor, na classe comercial. O grau de achatamento para as três classes aponta para curvas platicúrticas. Já a assimetria desta variável pode ser descrita como simétrica nas classes residencial e industrial e assimétrica moderada a esquerda (negativa) na classe comercial.

Com relação a variável da tarifa de energia elétrica (Tab. 4.15), o valor máximo está presente na classe comercial, enquanto que o valor mínimo está na classe industrial. A maior variabilidade está presente na classe industrial. Ainda, tem-se a indicação do grau de achatamento como sendo platicúrtica nas três classes. E com relação a assimetria, apontam-se que todas as três classes sejam consideradas assimétrica a direita (positiva).

A variável temperatura apresenta um baixo valor de desvio padrão, podendo ser considerada simétrica, apresentando os valores de média e mediana bem próximos. Com relação ao grau de achatamento da curva tem-se evidências de uma curva platicúrtica.

Para confirmação da sazonalidade, aplicou-se o teste do ANOVA com blocos sendo o resultado apresentado na Tab. 4.16. Para a variável do número de consumidores das classes comercial e industrial não rejeitam H_0 , logo não tem evidências de uma diferença significativa entre os meses. Com relação as demais variáveis, deve-se rejeitar a H_0 , encontrando evidências de uma diferença significativa entre os tratamentos (meses), bem como entre os blocos (anos).

Tabela 4.15 - Medidas de resumo dos dados mensais das variáveis exógenas para as três classes consumidoras/RS, 2004 - 19

Variável	Parâmetro	Residencial	Comercial	Industrial
NC	Valor máximo	4.114.961	351.044	56.542
	Valor mínimo	2.756.426	240.337	32.899
	Média	3.386.651	321.976	42.640
	Mediana	3.372.318	338.977	37.220
	Desvio padrão	392.510,8	26.810,74	8.541,66
	Coef. de variação (%)	11,6	8,3	20,0
	Coef. de assimetria	0,1	-0,6	0,5
	Coef. de curtose	-1,2	-1,0	-1,4
Tar	Valor máximo (R\$/MWh)	555,50	555,80	505,90
	Valor mínimo (R\$/MWh)	256,50	216,00	136,70
	Média (R\$/MWh)	345,60	325,10	277,60
	Mediana (R\$/MWh)	294,30	274,60	231,50
	Desvio padrão (R\$/MWh)	92,6	101,2	103,7
	Coef. de variação (%)	26,8	31,1	37,34
	Coef. de assimetria	0,95	0,94	0,77
	Coef. de curtose	-0,7	-0,7	-0,8
Temp	Valor máximo (°C)		25,82	
	Valor mínimo (°C)		10,75	
	Média (°C)		19,02	
	Mediana (°C)		19,23	
	Desvio padrão (°C)		3,99	
	Coef. de variação (%)		21,01	
	Coef. de assimetria		-0,19	
	Coef. de curtose		-1,22	

Obs: NC = número de consumidores; Tar = tarifa; Temp = Temperatura.

Fonte: Elaborada pela autora.

Tabela 4.16 - Resultado da ANOVA com blocos das variáveis externas para as três classes consumidoras/RS, 2004 - 19

Variável	Medidas	Fonte de variação		
		Blocos	Tratamentos	Erros
NC _{res}	gl	15	11	165
	f	13059,47	51,56	
	pvalor	<0,0001*	<0,0001*	
Tar _{res}	gl	15	11	165
	f	243,79	2,83	
	pvalor	<0,0001*	0,002	
NC _{com}	gl	15	11	165
	f	782,83	1,53	
	pvalor	<0,0001*	0,124	
Tar _{com}	gl	15	11	165
	f	314,42	5,33	
	pvalor	<0,0001*	<0,0001*	
NC _{ind}	gl	15	11	165
	f	239,60	1,71	
	pvalor	<0,0001*	0,076	
Tar _{ind}	gl	15	11	165
	f	363,30	6,02	
	pvalor	<0,0001*	<0,0001*	
Temp	gl	15	11	165
	f	2,08	176,73	
	pvalor	0,013*	<0,0001*	

Obs: * valor significativo para nível de significância de 5,0%.

Fonte: Elaborada pela autora.

A Tab. 4.17 apresenta o resultado do p-valor para a análise de resíduos dos testes ANOVA com bloco. Para todas as classes analisadas, com exceção do número de consumidores comerciais e da temperatura, rejeitou-se H_0 para o teste de autocorrelação, ou seja, os resíduos apresentam evidências de estarem autocorrelacionados. Com relação ao teste de normalidade a H_0 foi rejeitada para todas as variáveis, ou seja, foram encontradas evidências de que os resíduos não apresentam uma distribuição normal. Seguindo, no teste de homocedasticidade não se rejeita a H_0 , não existindo evidência de resíduos heterocedásticos, apenas para o número de consumidores residenciais e temperatura.

Tabela 4.17 - Análise dos resíduos do teste ANOVA com blocos das variáveis externas para as três classes consumidoras/RS, 2004 - 19

Variável	Ljung-Box (p-valor)	ARCH (p-valor)	Jarque-Bera (p-valor)
NC _{res}	0,0032*	1,000	<0,0001*
Tar _{res}	<0,0001*	0,0005*	<0,0001*
NC _{com}	0,1443	0,0017*	<0,0001*
Tar _{com}	<0,0001*	0,0004*	<0,0001*
NC _{ind}	<0,0001*	<0,0001*	<0,0001*
Tar _{ind}	<0,0001*	0,0006*	<0,0001*
Temp	0,1335	0,6499	0,0268*

Obs: * valor significativo para nível de significância de 5,0%.

Fonte: Elaborada pela autora.

Desta forma, o uso da ANOVA em blocos para verificação da sazonalidade não é válido, sendo aplicado o teste não paramétrico de Friedman, cujos resultados estão apresentados na Tab. 4.18.

Tabela 4.18 - Resultado do teste de Friedman das variáveis externas das três classes consumidoras/RS, 2004 - 19

Variável	Q	gl	p-valor
NC _{res}	167,46	11	<0,0001*
Tar _{res}	22,11	11	0,0236*
NC _{com}	42,45	11	<0,0001*
Tar _{com}	83,69	11	<0,0001*
NC _{ind}	7,25	11	0,7785
Tar _{ind}	96,12	11	<0,0001*
Temp	162,78	11	<0,0001*

Obs: * valor significativo para nível de significância de 5,0%.

Fonte: Elaborada pela autora.

Para todas as variáveis analisadas, com exceção do número de consumidores industrial, rejeita-se a hipótese nula, ou seja, revelam-se evidências de que em pelo menos um dos meses a distribuição da variável analisada é diferente, o que reforça a ideia de presença de sazonalidade nas séries.

No Apêndice 1 consta o diagrama de dispersão entre a demanda mensal de energia elétrica para cada uma das três classes consumidoras com as variáveis externas, em que foi feita uma análise preliminar da relação entre essas variáveis. Na sequência, são apresentados os resultados da análise de correlação cruzada entre a demanda mensal de energia elétrica e as variáveis externas.

As figuras 4.30, 4.31 e 4.32 mostram os gráficos resultantes desta análise da demanda mensal de energia elétrica da classe residencial com o número de consumidores, tarifa e temperatura, respectivamente.

O resultado desta análise em relação ao número de consumidores mostra a existência de uma correlação significativa na defasagem -8 na parte sazonal (Fig. 4.30(b)). Para variável tarifa foram detectadas correlações significativas até a décima sexta defasagem na parte sazonal (Fig. 4.31(b)). Com relação a temperatura, tem-se algumas correlações significativas até a defasagem -8 (Fig. 4.32(a)).

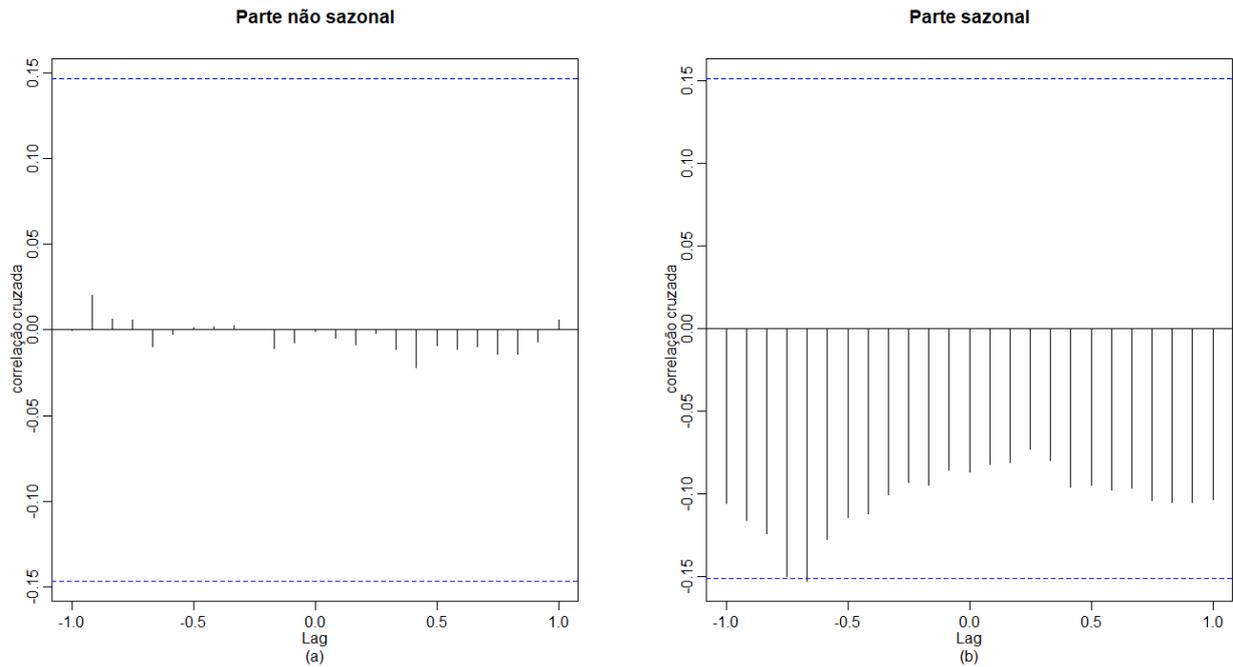


Figura 4.30 - Gráfico da correlação cruzada da parte não sazonal (a) e da parte sazonal (b) da demanda mensal de energia elétrica com o número de consumidores residenciais/RS, 2004 - 18

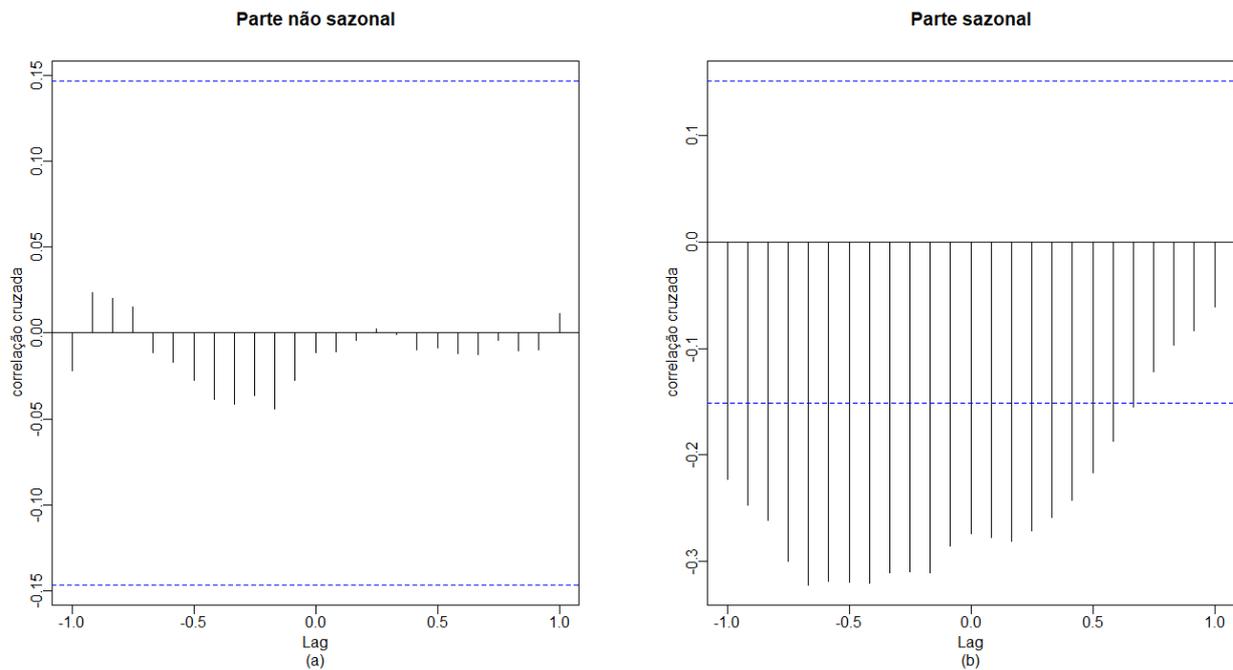


Figura 4.31 - Gráfico da correlação cruzada da parte não sazonal (a) e da parte sazonal (b) da demanda mensal de energia elétrica com a tarifa residencial/RS, 2004 - 18

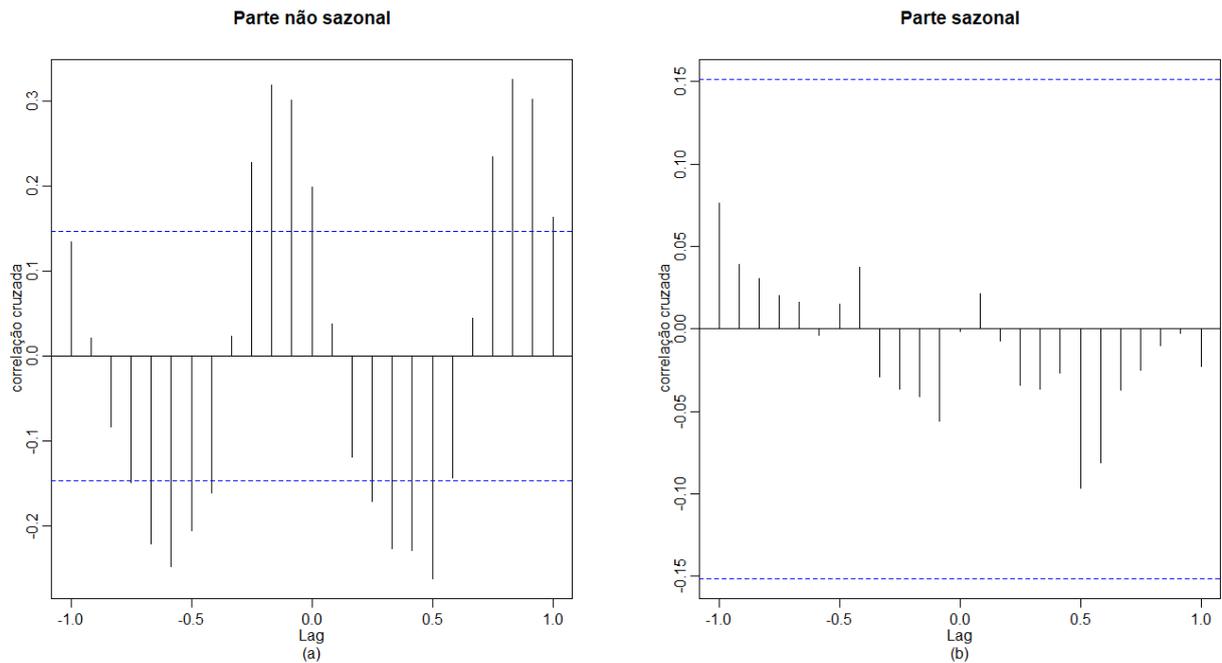


Figura 4.32 - Gráfico da correlação cruzada da parte não sazonal (a) e da parte sazonal (b) da demanda mensal de energia elétrica da classe residencial com a temperatura/RS, 2004 - 18

As Fig. 4.33, 4.34 e 4.35 mostram o resultado da correlação cruzada da demanda mensal de energia elétrica da classe comercial com o número de consumidores, tarifa e temperatura, respectivamente. Nesta análise, foram encontradas correlações significativas apenas com a temperatura, até a defasagem -10 para parte não sazonal, conforme mostra a Fig. 4.35(a).

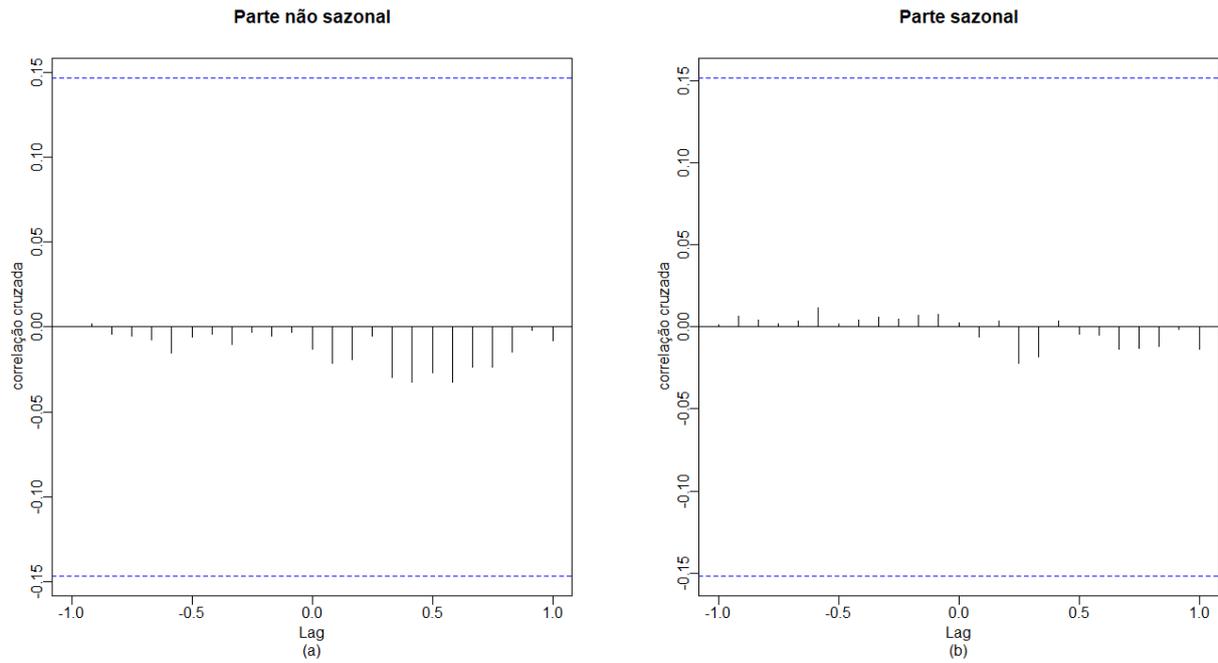


Figura 4.33 - Gráfico da correlação cruzada da parte não sazonal (a) e da parte sazonal (b) da demanda mensal de energia elétrica com o número de consumidores comerciais/RS, 2004 – 18

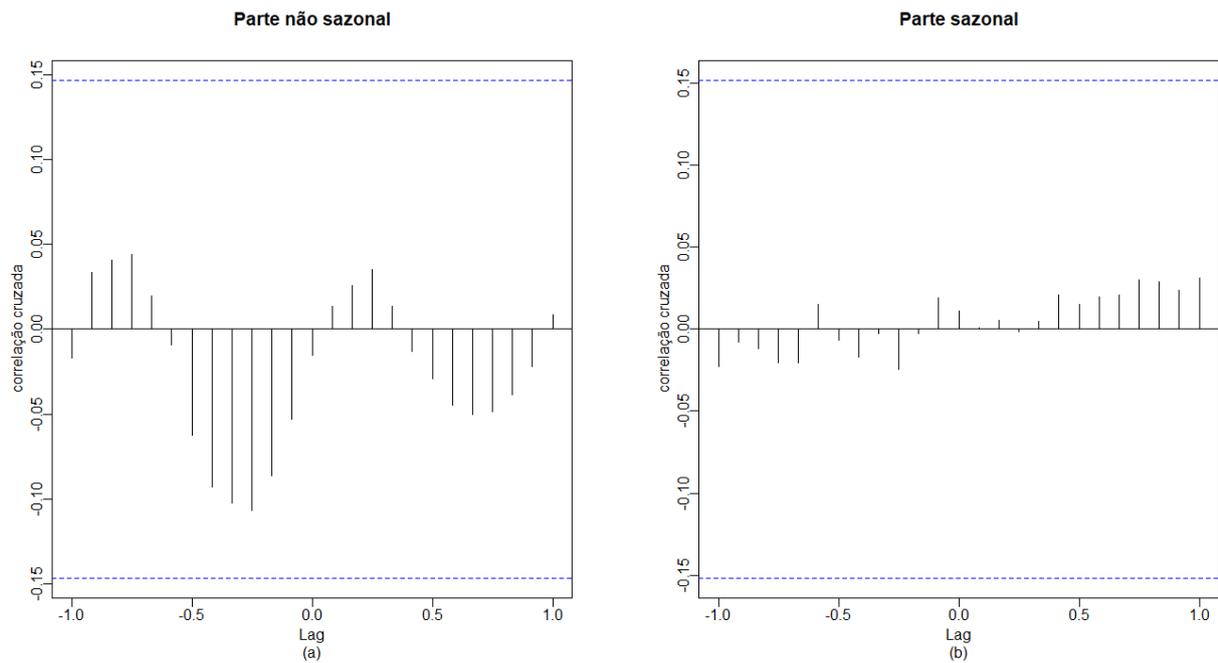


Figura 4.34- Gráfico da correlação cruzada da parte não sazonal (a) e da parte sazonal (b) da demanda mensal de energia elétrica com a tarifa comercial/RS, 2004 – 18

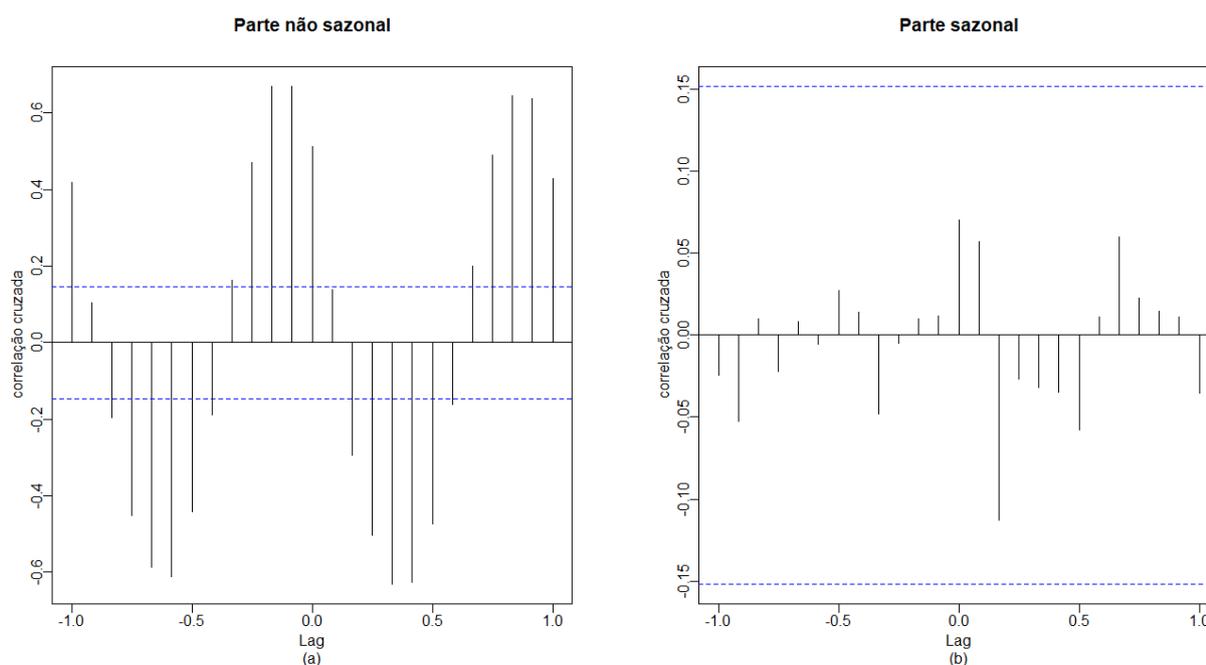


Figura 4.35- Gráfico da correlação cruzada da parte não sazonal (a) e da parte sazonal (b) da demanda mensal de energia elétrica da classe comercial com a temperatura/RS, 2004 – 18

As Fig. 4.36, 4.37 e 4.38 mostram o resultado da correlação cruzada da demanda mensal de energia elétrica da classe industrial com o número de consumidores, tarifa e temperatura, respectivamente. Para a variável número de consumidores tem-se uma correlação significativa na parte sazonal até a defasagem -6, conforme Fig. 4.36(b). Com a variável tarifa, também se nota correlação significativa na parte sazonal até a defasagem -5, conforme Fig. 4.37(b). Já com relação a variável temperatura, a correlação significativa aparece na parte não sazonal (Fig. 4.38(a)) em algumas defasagens até a -10.

Constatou-se que a associação das variáveis externas com a demanda mensal de energia elétrica industrial, foi bastante similar à associação destas variáveis com a demanda mensal de energia elétrica residencial, mas diferente da associação com a demanda mensal de energia elétrica comercial.

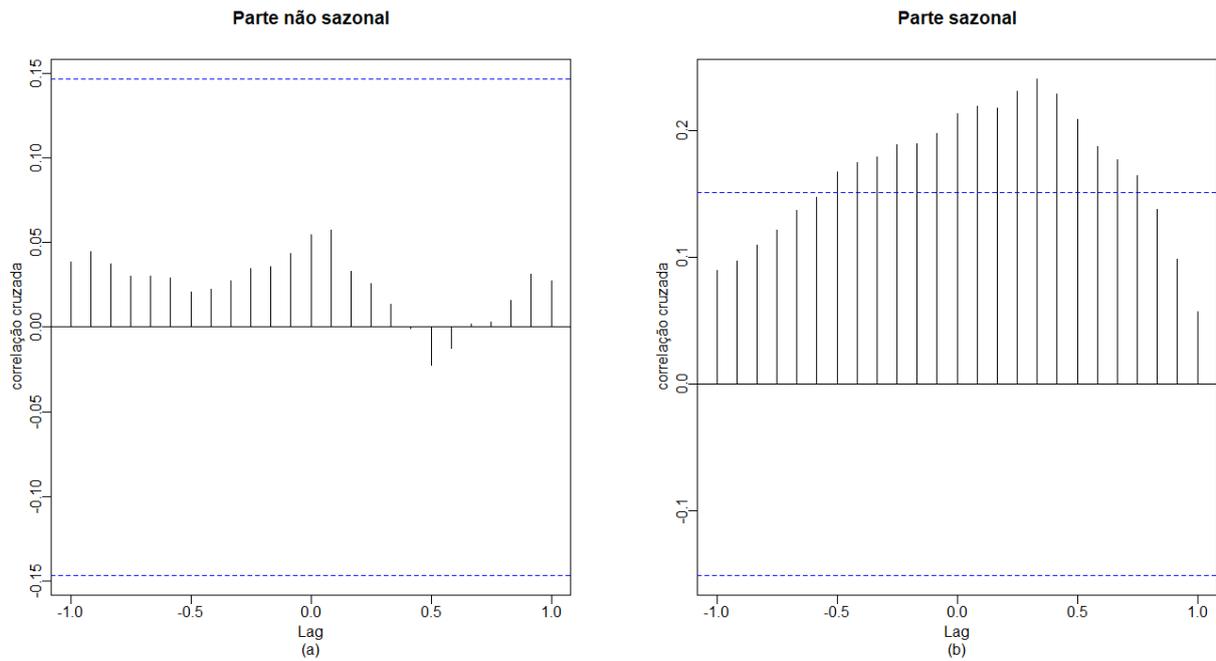


Figura 4.36 - Gráfico da correlação cruzada da parte não sazonal (a) e da parte sazonal (b) da demanda mensal de energia elétrica com o número de consumidores industriais/RS, 2004 – 18

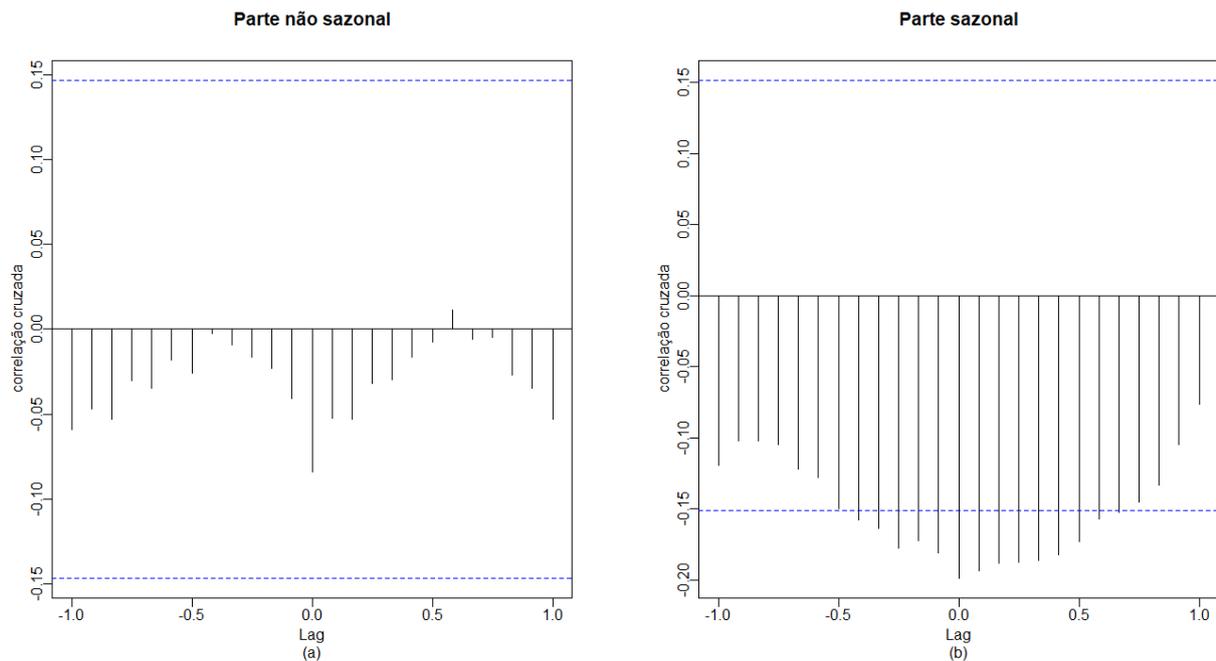


Figura 4.37 - Gráfico da correlação cruzada da parte não sazonal (a) e da parte sazonal (b) da demanda mensal de energia elétrica com a tarifa industrial/RS, 2004 – 18

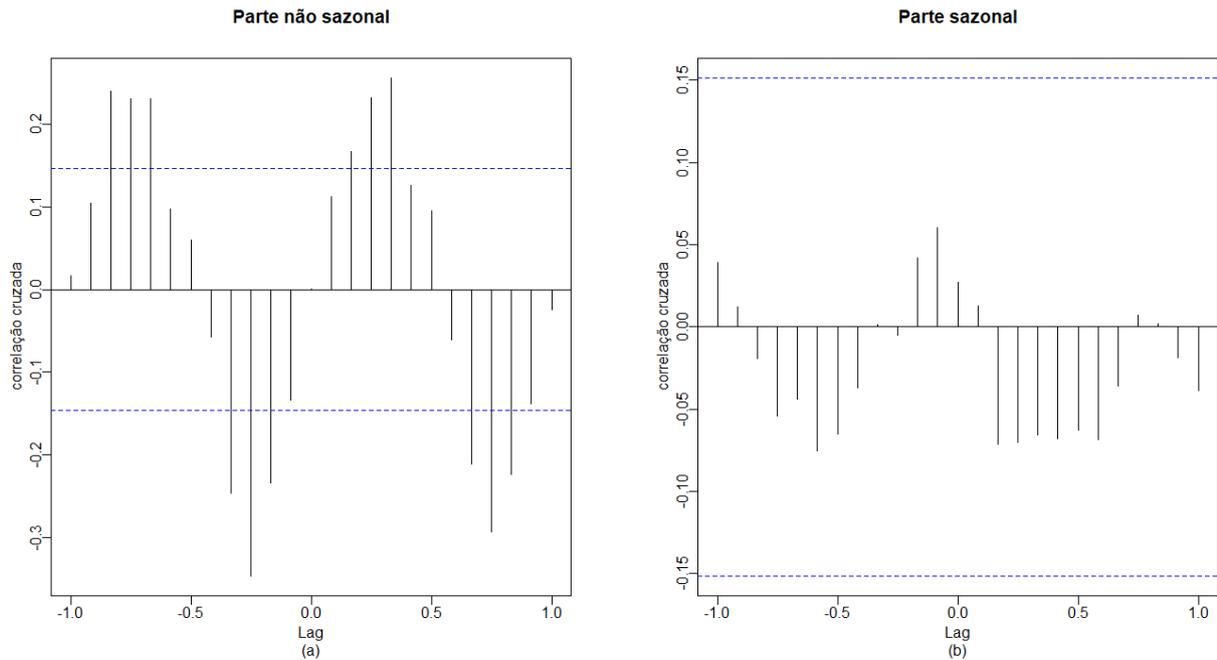


Figura 4.38 - Gráfico da correlação cruzada da parte não sazonal (a) e da parte sazonal (b) da demanda mensal de energia elétrica da classe industrial com a temperatura/RS, 2004 – 18

Com a determinação da quantidade de defasagens máximas a serem consideradas nas variáveis externas, estas foram inseridas aos modelos SARIMA já selecionados para as três classes consumidoras, avaliando a possibilidade de inserção desta quantidade máxima até uma defasagem, sendo posteriormente avaliadas as suas parcimonialidades através do BIC.

A Tab. 4.19 mostra os resultados do critério de informação do modelo residencial SARIMA $(1,1,2)(1,0,1)_{12}$ com cada uma das variáveis exógenas. Os modelos que apresentaram menor valor de BIC são SARIMAX $(1,1,2)(1,0,1)_{12}NC_{res_1}$, SARIMAX $(1,1,2)(1,0,1)_{12}Temp_1$ e SARIMAX $(1,1,2)(1,0,1)_{12}Tar_{res_1}$. A partir destes resultados foram analisadas todas as combinações possíveis, conforme mostra a Tab. 4.20.

Tabela 4.19 – Resultado do critério de informação do modelo residencial SARIMA (1,1,2)(1,0,1)₁₂ com cada uma das variáveis exógenas/RS, 2004 - 18

Defasagem máxima	BIC		
	Nº de consumidores	Tarifa	Temperatura
12		-512,29	
11		-521,72	
10		-529,84	
9		-539,17	
8	-546,43	-542,77	
7	-553,69	-550,43	
6	-563,07	-560,01	
5	-572,29	-569,16	
4	-575,76	-575,10	
3	-579,35	-584,57	-583,14
2	-588,03	-593,30	-591,34
1	-597,29*	-601,49*	-600,81*

Obs: * significa menor valor dentro da variável externa analisada.

Fonte: Elaborada pela autora.

Tabela 4.20 – Resultado do critério de informação do modelo residencial SARIMA (1,1,2)(1,0,1)₁₂ com combinações das variáveis exógenas/RS, 2004 - 18

Variáveis e defasagem máxima	BIC
NC _{res_1} , Tar _{res_1} , Temp_1	-611,00*
NC _{res_1} , Temp_1	-600,40
Tar _{res_1} , Temp_1	-602,73
NC _{res_1} , Tar _{res_1}	-606,28

Obs: * significa menor valor.

Fonte: Elaborada pela autora.

A Tab. 4.21 apresenta os resultados do BIC para o modelo da demanda mensal de energia elétrica comercial SARIMA (1,1,2)(0,1,1)₁₂ com todas as defasagens da variável temperatura, a única com a qual foram encontradas evidências de correlação.

Tabela 4.21- Resultado do critério de informação do modelo comercial SARIMA (1,1,2)(0,1,1)₁₂ com cada umas das variáveis exógenas/RS, 2004 - 18

Defasagem	Variável	BIC
10	Temperatura	-560,42
9	Temperatura	-568,34
8	Temperatura	-577,21
7	Temperatura	-587,39
6	Temperatura	-593,70
5	Temperatura	-600,81
4	Temperatura	-610,88
3	Temperatura	-619,35
2	Temperatura	-628,34
1	Temperatura	-638,33*

Obs: * significa menor valor.

Fonte: Elaborada pela autora.

A Tab. 4.22 mostra o resultado do critério de informação do modelo industrial SARIMA (1,1,1)(1,0,2)₁₂ com cada umas das variáveis externas e suas respectivas defasagens. Os modelos mais parcimoniosos são: o SARIMAX (1,1,1)(1,0,2)₁₂NC_{ind_3}, SARIMAX (1,1,1)(1,0,2)₁₂Tar_{ind_2}, SARIMAX (1,1,1)(1,0,2)₁₂Temp₂. Com isto, selecionou-se estas variáveis externas, com suas respectivas defasagens, sendo feitas modelagens com as combinações possíveis, conforme mostra a Tab. 4.23.

Diferente da classe residencial, em que a combinação com mais de uma variável apresentou modelos mais parcimoniosos, na classe industrial, o modelo com melhor resultado foi o SARIMAX (1,1,1)(1,0,2)₁₂NC_{ind_3}, que considera apenas a variável do número de consumidores.

Tabela 4.22 - Resultado do critério de informação do modelo industrial SARIMA (1,1,1)(1,0,2)₁₂ com cada umas das variáveis exógenas/RS, 2004 - 18

Defasagem	BIC		
	Nº de consumidores	Tarifa	Temperatura
6	-547,39		
5	-554,63	-545,67	
4	-563,21	-553,35	-552,73
3	-567,61*	-559,73	-558,31
2	-563,66	-564,03*	-565,13*
1	-562,96	-560,63	-562,22

Obs: *significa menor valor dentro da variável externa analisada.

Fonte: Elaborada pela autora.

Tabela 4.23 - Resultado do critério de informação do modelo industrial SARIMA (1,1,1)(1,0,2)₁₂ e a combinação das variáveis exógenas/RS, 2004 - 18

Variáveis e defasagem máxima inseridas	BIC
NC _{ind_3} , Tar _{ind_2} , Temp_2	-558,85
NC _{ind_3} , Temp_2	-563,04
Tar _{ind_2} , Temp_2	-558,84
NC _{ind_3} , Tar _{ind_2}	-563,28*

Obs: * significa menor valor.

Fonte: Elaborada pela autora.

Para análise dos resíduos foram selecionados todos os modelos que apresentaram valores do BIC menor em relação ao BIC do modelo SARIMA da sua respectiva classe. Além disto, também foram analisados os resíduos dos modelos de menor BIC quando estes incluíam uma única variável externa, mesmo que o seu valor apresentado fosse maior em relação ao BIC do modelo SARIMA da sua respectiva classe. A Tab. 4.24 apresenta o resultado da análise dos resíduos para as três classes consumidoras.

Tabela 4.24 - Resultado da análise dos resíduos dos modelos SARIMAX nas três classes consumidoras/RS, 2004 - 18

Classe	Modelo	Variável externa	Ljung-Box (p-valor)	Jarque-Bera (p-valor)	ARCH (p-valor)
Res	SARIMA (1,1,2)(1,0,1) ₁₂	NC _{res_1}	0,0092*	0,0001*	0,0502
		Tar _{res_1}	0,0138	0,0008*	0,0188
		Temp_1	0,0073*	0,0039*	0,0857
		NC _{res_1} , Tar _{res_1} , Temp_1	0,0061*	0,0005*	0,0056*
		NC _{res_1} , Tar _{res_1}	0,0111	<0,0001*	0,0005*
Com	SARIMA (1,1,2)(0,1,1) ₁₂	Temp_3	<0,0001*	0,1654	0,4981
		Temp_2	0,0004*	0,1580	0,4228
		Temp_1	0,0005*	0,1580	0,4033
Ind	SARIMA (1,1,1)(1,0,2) ₁₂	NC _{ind_3}	0,9997	<0,0001*	0,9981
		Tar _{ind_2}	0,9859	<0,0001*	0,9376
		Temp_2	0,9925	<0,0001*	0,8847
		NC _{ind_3} , Tar _{ind_2}	0,9993	<0,0001*	0,9992

Obs: * valor significativo para nível de significância de 1,0%.

Fonte: Elaborada pela autora.

Nota-se na classe residencial que apenas dois modelos, SARIMA (1,1,2)(1,0,1)₁₂Tar_{res_1} e SARIMA (1,1,2)(1,0,1)₁₂NC_{res_1},Tar_{res_1}, não permitem rejeitar H₀ para o teste de independência e, destes dois modelos, apenas o modelo SARIMA (1,1,2)(1,0,1)₁₂Tar_{res_1} não rejeita H₀ para o teste de homocedasticidade. Já para o teste de normalidade todos os modelos apresentam evidências de não

apresentarem resíduos normalmente distribuídos. Na classe comercial todos os modelos permitem rejeitar H_0 para o teste de autocorrelação, logo seus resíduos não são independentes e não é possível criar modelos de previsão através da técnica SARIMAX com os critérios adotados. Já na classe industrial em todos os modelos não se rejeita H_0 para o teste de Ljung-Box e ARCH, ou seja, pode-se considerar que os resíduos não apresentam autocorrelação e nem heterocedasticidade, porém o mesmo não pode ser considerado sobre normalidade.

Dos modelos apresentados e avaliados na Tab. 4.24, foram selecionados aqueles que não apresentam resíduos autocorrelacionados para analisar sua capacidade preditiva. A Tab. 4.25 apresenta as medidas de acurácia: RMSE, MAE, MAPE e o U de Theil dos modelos selecionados. Nota-se que, para a classe residencial, os menores valores para estes indicadores foram obtidos o modelo SARIMA $(1,1,2)(1,0,1)_{12}$ Tar_{res_1}, e para classe industrial, para o modelo SARIMA $(1,1,1)(1,0,2)_{12}$ NC_{ind_3}, Tar_{ind_2}.

Tabela 4.25 - Indicadores de qualidade para validação dos modelos SARIMAX das três classes/RS, 2019

Classe	Modelo	Variável	RMSE	MAE	MAPE(%)	U de Theil
Res	SARIMA $(1,1,2)(1,0,1)_{12}$	Tar _{res_1}	42675,56*	33811,74*	4,3308*	0,5475*
		Temp_1, NC _{res_1}	46464,77	37218,23	4,9532	0,6161
Ind	SARIMA $(1,1,1)(1,0,2)_{12}$	NC _{ind_3}	34823,90	29518,94	3,6499	1,0539
		Tar _{ind_2}	31392,06	23436,22	2,9271	0,9727
		Temp_2	37710,10	31450,03	3,9071	1,1623
		NC _{ind_3} , Tar _{ind_2}	28528,50*	21281,82*	2,6505*	0,8744*

Obs: * significa menor valor.

Fonte: Elaborada pela autora.

A tabelas 4.26 e 4.27 mostram os valores observados, previsto e os intervalos de confiança para o modelo SARIMAX das classes residencial e industrial, respectivamente. Nota-se que na classe residencial existem dois valores observados fora do intervalo de confiança e na classe industrial apenas um valor fora do intervalo.

Tabela 4.26 - Valores observados, previstos e intervalo de confiança do modelo residencial SARIMA (1,1,2)(1,0,1)₁₂ Tar_{res}_1/RS, 2019

Meses/2019	Valor observado	Previsão	IC-95%	
Janeiro	917352*	820712,38	763663,10	882023,52
Fevereiro	881917	827267,19	763721,90	896099,74
Março	813091	816168,41	753474,91	884077,49
Abril	692676	743367,64	686266,28	805219,35
Maiο	712893	735330,25	678846,96	796513,22
Junho	666041*	747801,61	690359,67	810022,25
Julho	718333	771271,71	712027,63	835446,01
Agosto	725017	742712,27	685661,95	804509,46
Setembro	683906	733208,95	676887,93	794215,41
Outubro	726320	706737,08	652450,14	765541,73
Novembro	735331	734073,91	677687,13	795152,34
Dezembro	772078	776397,57	716759,04	840997,53

Obs: *valor observado fora do IC-95.

Fonte: Elaborada pela autora.

Tabela 4.27 - Valores observados, previstos e intervalo de confiança do modelo industrial SARIMA $(1,1,1)(1,0,2)_{12}NC_{ind_3}, Tar_{ind_2}/ RS, 2019$

Meses/2019	Valor observado	Previsão	IC-95%	
Janeiro	744672*	679842,2	629758,7	733908,8
Fevereiro	797519	771694,5	706099,9	843381,8
Março	812329	811939,4	738633,4	892520,7
Abril	828741	824001,0	746651,6	909362,4
Maiο	847838	809388,3	731010,4	896168,7
Junho	848682	812591,7	731716,2	902407,1
Julho	800584	828742,7	744149,3	922951,5
Agosto	848966	846647,7	758154,1	945471,5
Setembro	825402	819923,2	732276,9	918060,9
Outubro	801219	808717,5	720401,5	907861,4
Novembro	831736	797834,1	708910,1	897912,6
Dezembro	779373	771671,3	683966,4	870621,8

Obs: * valor observado fora do IC-95.

Fonte: Elaborada pela autora.

As figuras 4.39 e 4.40 mostram os gráficos que representam os modelos SARIMA $(1,1,2)(1,0,1)_{12} Tar_{res_1}$ e SARIMA $(1,1,1)(1,0,2)_{12}NC_{ind_3}, Tar_{ind_2}$, identificados como aqueles com melhor capacidade preditiva para as classes residencial e industrial, respectivamente. É possível perceber que no período de validação, os valores observados encontram-se dentro dos intervalos de confiança quase que em sua totalidade, indicando que os modelos são satisfatórios.

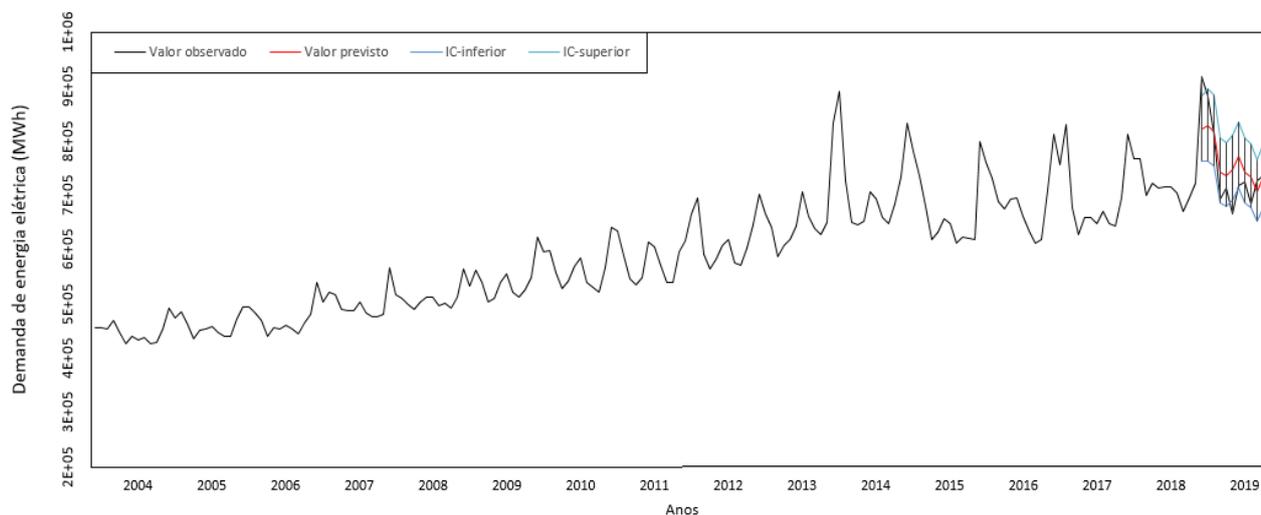


Figura 4.39 - Gráfico do modelo de previsão SARIMA $(1,1,2)(1,0,1)_{12}$ Tar_{res_1} para demanda mensal de energia elétrica da classe residencial/RS, 2004 - 19

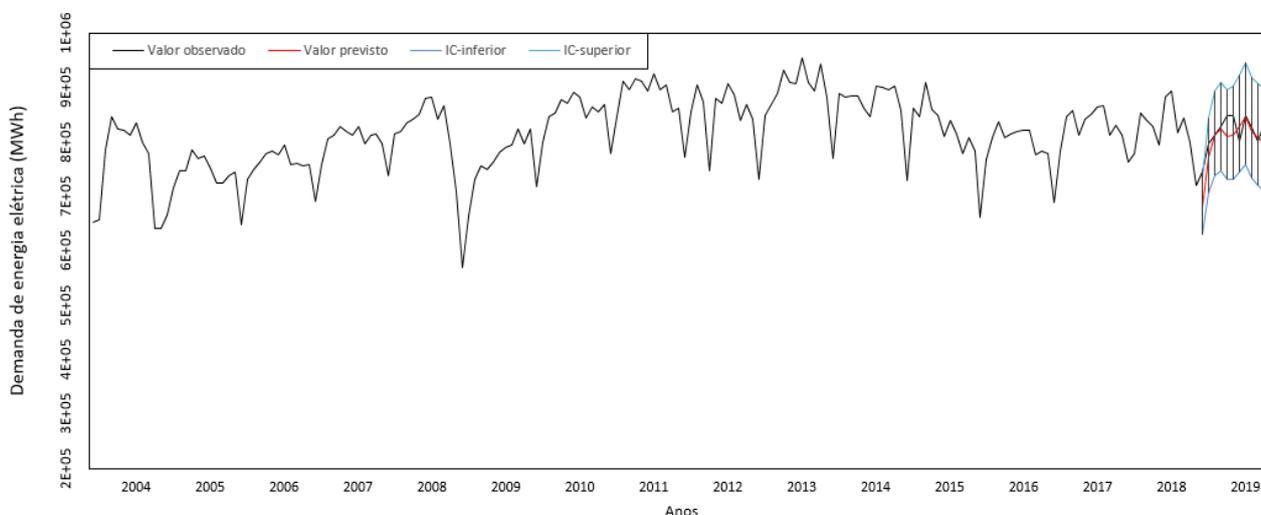


Figura 4.40 - Gráfico do modelo de previsão SARIMA $(1,1,1)(1,0,2)_{12}$ NC_{ind_3}, Tar_{ind_2} para demanda mensal de energia elétrica da classe industrial/RS, 2004 - 19

Com relação aos resultados apresentados nesta seção pode-se verificar que quanto menos defasagens é inserida a variável externa na modelagem, em geral, mais parcimoniosos são os modelos desenvolvidos.

Para classe residencial o modelo escolhido não foi o mais parcimonioso e considera apenas uma variável externa, tarifa com 1 defasagem, porém seus resíduos apresentam evidências de não estarem autocorrelacionados e serem homocedásticos. Além disso, seus valores para os indicadores de qualidade foram melhores que os dos outros modelos analisados.

A classe comercial foi a única que apresentou evidências de resíduos não independentes e, por isso, todos os modelos analisados, conforme critérios adotados, foram descartados.

Assim como na classe residencial, a classe industrial obteve um modelo de previsão que não é o mais parcimonioso, considerando duas variáveis externas: número de consumidores e tarifa, com 3 e 2 defasagens, respectivamente. Com relação a análise dos resíduos, este modelo não satisfaz apenas a suposição de normalidade. Já sobre os indicadores de qualidade, entre os modelos avaliados, foi o que apresentou os menores valores.

4.3. Avaliação dos modelos desenvolvidos, métodos e técnicas utilizadas

Analisando o resultado apresentado pelos dois modelos desenvolvidos na classe residencial tem-se que o modelo SARIMA $(1,1,2)(1,0,1)_{12}$ é mais parcimonioso que o SARIMA $(1,1,2)(1,0,1)_{12}$ Tar_{res_1}, conforme apresentado pelas tabelas 4.6, 4.19 e 4.20. Na análise dos resíduos, tabelas 4.6 e 4.24, ambos os modelos falharam apenas no teste de normalidade. Embora os valores dos indicadores de qualidade dos dois modelos foram muito próximos, tabelas 4.9 e 4.25, o modelo SARIMA $(1,1,2)(1,0,1)_{12}$ apresentou os menores valores. Com relação aos valores observados fora do intervalo de confiança, no modelo SARIMA $(1,1,2)(1,0,1)_{12}$, Tab. 4.12, apresentou-se apenas um, enquanto que no modelo SARIMA $(1,1,2)(1,0,1)_{12}$ Tar_{res_1}, Tab. 4.26, foram observados dois valores. Com isso, o melhor modelo para a classe residencial é o SARIMA $(1,1,2)(1,0,1)_{12}$ sem a inclusão da variável externa.

A classe industrial apresentou um melhor valor de parcimonia no modelo SARIMA $(1,1,1)(1,0,2)_{12}$, conforme verifica-se nas tabelas 4.8, 4.22 e 4.23. A análise dos resíduos, tabelas 4.8 e 4.24, é semelhante à residencial, ou seja, apenas não apresentaram evidências de normalidade. Com relação aos indicadores de qualidade, tabelas 4.11 e 4.25, o modelo SARIMA $(1,1,1)(1,0,2)_{12}$ NC_{ind_3}, Tar_{ind_2} apresentou menores valores, logo este é o melhor modelo a ser utilizado na previsão desta série. No tocante aos valores observados fora do intervalo de confiança ambos os modelos apresentam apenas um valor fora do limite aceitável, conforme mostra as tabelas 4.14 e 4.27.

Com relação aos métodos e técnicas utilizados, foram identificadas algumas fragilidades. A análise da presença de sazonalidade foi prevista pelos gráficos em linhas e subséries, sendo complementada em concordância com os resultados do teste ANOVA com blocos e Friedman.

Para análise de estacionariedade foi empregado o teste ADF, o que foi feito separadamente para a parte não sazonal e para a parte sazonal. Seu desempenho para a parte não sazonal foi considerado satisfatório por ter acordado com o teste KPSS, o mesmo não acontecendo em relação à parte sazonal. Neste caso, quando comparado com os resultados dos testes OSBD e HEDY, houve

discordância. Logo, os testes para analisar a estacionariedade necessitam ser mais estudados para facilitar as tomadas de decisões.

O método da máxima verossimilhança utilizado na estimação dos coeficientes dos modelos, executado com a função “*arima*” do pacote *forecast* do *software R*, apresentou algumas limitações. Alguns dos 256 modelos considerados inicialmente não conseguiram ser estimados, mesmo utilizando os métodos de otimização disponibilizados. Em geral, nas classes residencial e industrial estes modelos apresentam $P=3$ e na classe comercial $p=3$. É importante ressaltar que o uso da função *auto.arima*, resultou em modelos com valores de p , P , q e Q entre 1 e 2, assim houve uma concordância em não estimar modelos com ordem maiores.

Outro aspecto a considerar diz respeito à dinâmica de inserção das variáveis externas nos modelos SARIMAX: estas foram inseridas em modelo já definido, sendo as variáveis e suas quantidades de defasagens definidas pela correlação cruzada. Poderia ter sido estudado primeiro a relação entre as variáveis externas e a variável dependente usando causalidade de Granger e correlação cruzada para depois realizar a modelagem SARIMAX já com a inserção das variáveis externas. Outra possibilidade consistiria em fazer trabalho conjunto com regressão múltipla e SARIMAX como feito em Tarsitano e Amerise (2017).

5. CONSIDERAÇÕES FINAIS

O presente trabalho teve como objetivo analisar o comportamento da sazonalidade na demanda mensal de energia elétrica do estado do Rio Grande do Sul para as três principais classes consumidoras: residencial, comercial e industrial. A principal motivação para o desenvolvimento deste trabalho é tentar se aproximar dos padrões e comportamentos dessa variável e assim, reduzir os erros de previsão, auxiliando diretamente nas decisões técnicas e financeiras para este setor.

Para isso, uma análise exploratória, com técnicas gráficas e analíticas, foi desenvolvida e constatou-se a presença de sazonalidade nas três classes. Além disso, as análises gráficas, mostraram que as classes residencial e comercial apresentam comportamentos bastante semelhantes, com meses de maior valor de suas demandas em janeiro, fevereiro e março, o que já não acontece na classe industrial.

Com a confirmação da presença de sazonalidade, utilizou a metodologia de Box-Jenkins para séries temporais com essa característica e modelou-se a demanda mensal de energia elétrica nas três classes. Empregou-se o modelo Autorregressivo Integrado de Média Móvel Sazonal (SARIMA) e com base no resultado desta modelagem fez-se um refinamento empregando a modelagem SARIMAX (modelagem SARIMA com fator exógeno), nesta etapa utilizou-se como variáveis externas o número de consumidores, o valor da tarifa de energia e temperatura.

Os resultados mostram que as classes residencial e comercial obtiveram melhor desempenho com o modelo SARIMA, sendo o modelo SARIMA(1,1,2)(1,0,1)₁₂ para a classe residencial e o modelo SARIMA(1,1,2)(0,1,1)₁₂ para classe comercial. Já para a classe industrial, o modelo SARIMAX, com a inserção de duas variáveis externas, SARIMA(1,1,1)(1,0,2)₁₂NC_{ind_3},Tar_{ind_2}, obteve os melhores valores na etapa de validação.

No decorrer do desenvolvimento deste trabalho, foram encontrados três pontos de fragilidade na metodologia utilizada. O primeiro é com relação ao uso do teste ADF e, em alguns casos, a discordância de resultados quando comparados a outros testes para verificação da estacionariedade, principalmente em relação à parte sazonal. O segundo é a limitação de não conseguir estimar os 256 modelos candidatos para as três classes analisada, mesmo empregando os métodos de otimização disponibilizados pela função “*arima*” do pacote “*forecast*” do *software* R. O terceiro, e último ponto, é a não normalidade constatada na análise dos resíduos para os modelos da classe residencial e industrial, além da presença de valores discrepantes conforme verificado nos gráficos de caixa.

Uma hipótese de ser avaliada em trabalhos futuros é buscar melhorar alguns dos pontos apresentados no parágrafo anterior. Além disso, é exequível a escolha e estudo de outras variáveis

para cada uma das três classes analisadas, bem como a utilização de outras dinâmicas de análise do seu relacionamento com a demanda de energia elétrica na modelagem SARIMAX. É importante ressaltar que as variáveis externas apresentaram ordens de integração diferentes, conforme apresentado no apêndice 2, e que estas não foram consideradas na modelagem.

Os resultados encontrados neste trabalho podem ser contrastados com os obtidos por Nunes (2019). Este autor realizou a modelagem de dados anuais da demanda de energia elétrica para as classes consumidoras residencial, comercial e industrial, do estado do Rio Grande do Sul, através das metodologias Box-Jenkins, Vetores autoregressivos (VAR) e Vetores Autoregressivos em conjunto com Correção de Erros Vetoriais (VAR-VEC), sendo que estes dois últimos modelos empregam variáveis externas e, dentre as variáveis empregadas pelo autor, a que coincide com as utilizadas neste trabalho é a tarifa. Nunes (2019) mostra que as classes comercial e industrial, apresentam melhor desempenho sem considerar as variáveis externas, modelo ARIMA. Já a classe residencial tem melhores previsões com a modelagem VAR. Em suma, a classe comercial, em ambos os trabalhos desenvolvidos, apresenta uma melhor previsão com modelos que não consideram a inserção de variáveis externas.

Por fim, vale realçar que os modelos encontrados podem ser utilizados para previsão da demanda mensal de energia elétrica no estado do Rio Grande do Sul, fornecendo subsídios para a gestão do processo de produção, transmissão e distribuição de energia elétrica, de maneira a minimizar perdas. Também podem ser usados para elaboração de políticas públicas relacionadas ao desenvolvimento econômico da região, tendo em vista sua proximidade como setor elétrico.

REFERÊNCIAS

- AKAIKE, Maximum likelihood identification of Gaussian autoregressive moving average models. **Biometrika**, v.60, n.2, p. 255-265, 1973.
- ANEEL, **Bandeiras Tarifárias**. Agência Nacional de Energia Elétrica. Brasília, 2015. Disponível em: <https://www.aneel.gov.br/bandeiras-tarifarias>. Acesso em: jan. 2020.
- ANEEL, **Tarifa Branca**. Agência Nacional de Energia Elétrica. Brasília, 2020. Disponível em: <https://www.aneel.gov.br/tarifa-branca>. Acesso em: jan. 2020.
- ARUNRAJ, N.S.; AHRENS, D. A. Hybrid Seasonal Autoregressive Integrated Moving Average and Quantile Regression for Daily Food Sales Forecasting, **International Journal of Production Economics**, v. 170, p,321-335, 2015.
- AYLAR, E. SMEEKES, S. WESTERLUND, J. Lag truncation and the local asymptotic distribution of the ADF test for a unit root, **Stat Papers**, v. 60, p, 2109 – 2118, 2019.
- BARBETTA, P. A., REIS M. M., BORNIA A. C. **Estatística: para Cursos de Engenharia e Informática**. São Paulo: Atlas S.A, 2004. 410p. ISBN: 8522437653.
- BUENO, R. D. L.S, **Econometria de séries temporais**. 2 ed, São Paulo: Editora Cengage Learning, 2011, 360p,
- CABRAL, J. DE A., LEGEY, L. F. L., FREITAS, M. V. DE. Electricity consumption forecasting in Brazil: A spatial econometrics approach. **Energy**, v.126, p.124-131, 2017.
- CRYER, J.D.; CHAN, K.; **Time Series Analysis Whit Applicatons in R**, New York: Spring, 2º ed, 2008.
- COWPERTWAIT, P. S. P e METCALFE, A. V. **Introductory Time Series with R**. p. 254, New York:Springer, 2009.
- DICKEY, D. A.; FULLER, W. A. Likelihood ratio statistics for autoregressive time series with a unit root. **Econometrica**, 49:1057-1073, 1981
- ELAMIN, N., FUKUSHIGE, M., Modeling and Forecasting Hourly Electricity Demand by SARIMAX with Interactions. **Energy**, v. 165, p. 257-268, 2018.
- EPE, Empresa de Pesquisa Energética, **Consumo Mensal de Energia Elétrica por classe (regiões e subsistemas)**, Rio de Janeiro, Brasil, 2020 Disponível em:<<https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/Consumo-mensal-de-energia-eletrica-por-classe-regioes-e-subsistemas>>. Acesso em: 3 de maio, de 2020.
- ELLIOTT, G., ROTHENBERG , T. J. e STOCK, J. H., Efficient Test For Na Autoregressive Unit Rott, **Econometrica**, v. 64, n. 4, p. 813 – 896, 1996.
- ESPINOSA, M.M. PRADO, S. M. GHELLER, M., Uso do modelo SARIMA na previsão do número

de focos de calor para os meses de junho a outubro no Estado de Mato Grosso, **Ciência e Natura**, UFSM, 32,(2):7-21,2010.

- FERREIRA, P,G,C, et al, **Análise de Séries Temporais em R: curso introdutório**. Elsevier, 2017.
- GRETl - **GNU Regression, Econometrics and Timeseries Library**, 2021. Disponível em: <http://gretl.sourceforge.net/>. Acesso em: jan. 2021.
- GUJARATI , D.N.; PORTER, D.C. **Econometria Básica**, Porto Alegre: Editora AMGH, 2011.
- HAMILTON, J. D. **Time series analysis**. Princeton, Princeton, 1993.
- HYNDMAN, R.J., & ATHANASOPOULOS, G. **Forecasting: principles and practice**, 2° ed, OTexts: Melbourne, Austrália, 2018. OTexts.com/fpp2. Acesso em: fev. 2020.
- HYLLEBERG, S., ENGLE, R.F., GRANGER , C.W.J. e YOO, B.S. Seasonal integration and cointegration, **Journal of Econometrics**, 44, 215-238, 1990.
- HURVITCH, C.M.; TSAI, C. Bias of the corrected aic criterion for underfitted regression and time series models. **Biometrika**, 78(3):449-509, 1991.
- INMET, Instituto Nacional de Meteorologia, **Dados históricos**, Brasília, 2020. Disponível em: <https://portal,inmet.gov.br/dadoshistoricos>. Acesso em: dez. 2020.
- IPEA. **Macroeconômico**, 2020. Disponível em: <http://www.ipeadata.gov.br/Default.aspx>. Acesso em: dez. 2020.
- 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.; PHILIPS, P. C.B.; SCHIMDT, P.; SCHIN, Y. 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.
- LEVENBACH, H.; **Change&Chance Embraced**, Editora: David Coen Freelance copyeditor, 2017.
- LI, M., ALLINSON, D. E HE, M. Seasonal variation in household electricity demand: A comparison of onitored and synthetic daily load profiles. **Energy & Buildings**, vol. 179, 292-300p., 2018.
- LIMA, L. M., BACCHI M. R. P. Assessing the impact of Brazilian economic growth on demand for electricity. **Energy**, v. 172, p. 861-873, 2019.
- LJUNG, G. M.; BOX, G. E. P. On a Measure of Lack of Fit in Time Series Models. **Biometrika**, 65(2):297-303, 1978.
- MORETTIN, P, A.,; TOLOI, C.M. C. **Análise de Séries Temporais Modelos lineares univariados**, 3° ed. São Paulo: Blücher, 474 p., 2018. ISBN: 978-85-212-1352-9.
- NUNES, G. S. **O uso dos métodos ARIMA e VAR-REC no estudo da demanda de energia elétrica**

no Rio Grande do Sul, 2019. Dissertação de mestrado, RS: Universidade Federal do Rio Grande (FURG).

OLIVEIRA, A. DE, SILVEIRA, G. B., BRAGA, J. DE M. **Diversidade Sazonal do Consumo de Energia Elétrica no Brasil**, 2001. Disponível em: <http://repositorio.ipea.gov.br/handle/11058/3904>. Acesso em: jan. 2020

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

PEREIRA, C. M. **Modelagem Fuzzy para Previsão de uma Série Temporal de Distribuição de Energia Elétrica**, 2015. Dissertação de Mestrado, RJ: Universidade do Estado do Rio de Janeiro (UERJ)

PHILIPS, P.; PERRON, P. Testing for a unit root in time series regression. **Biometrika**, 75(2):335-346, 1988.

PINTO, W.P, REISEN, V. A., MONTE, E. Z. Previsão da concentração de material particulado inalável, na Região Grande Vitória, ES, Brasil, utilizando o modelo SARIMAX. **Engenharia Sanitária e Ambiental**, Rio de Janeiro, v. 22, n. 2, p. 307 - 318, mar/abr. 2018. DOI: 10.1590/S1413-41522018168758

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:<<https://www.R-project.org/>,> . Acesso em: abr. 2020.

SCHEIDT, F; MEDINOVA, H; LUDWIG, N; RICHTER, B; STAUDT, P; WEINHARDT, C; Data analytics in the electricity sector –A quantitative and qualitative literature review. **Energy and AI**.

SIEGEL, S., JR, N. J. C., **Estatística não-paramétrica para ciências do comportamento**. 2º ed. Porto Alegre: Artmed, 2006. 448p. ISBN 8536307293

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

SCHWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, 6(2):461-464, 1978.

SOUZA, R. C., CAMARGO, M. E. **Análise e previsão de séries temporais: os modelos ARIMA**. Ijuí: SEDIGRAF, 1996.

TARSIANO, A., AMERISE, I.L., Short-term load forecasting using a two-stage sarimax model. **Energy**. v. 133. p. 108-114, 2017. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0360544217308848>>. Acesso: jun. 2020.

THEIL, H. **Applied economic forecasting**. Amsterdam: North-Holland, 1966

THOMAZ, R. A. **Demanda de energia elétrica no Brasil: 1995-2015**, 2017. Dissertação de Mestrado, PR: Universidade Federal do Paraná (UFPR).

THOMAZ, P. S. **Modelos ARMA-GARCH na modelagem da volatilidade de ações financeiras**, 2019. Dissertação de Mestrado, RS: Universidade Federal do Rio Grande (FURG).

VALER, L. A. **Modelo matemático ARIMAX de um propulsor eletromecânico utilizado em naves do tipo multirrotor**, 2016. Dissertação de mestrado, RS: Universidade Regional do Noroeste do Estado do Rio Grande do Sul (Unijuí).

U. S. Census Bureau, **X-13ARIMA-SEATS Reference Manual Accessible HTML Output Version**, United States Census Bureau, 2017 Disponível em: <https://www.census.gov/ts/x13as/docX13AS.pdf>.

ZIVOT, E. & ANDREWS, D. K., Further evidence on the great crash, the oil-price shock, and the unit-root, **Journal of Business and Economic Statistics** 10(3), 251–270, 1992.

APÊNDICE 1 – Diagramas de dispersão entre as demandas mensais de energia elétrica com as variáveis exógenas

As figuras 0.1, 0.2 e 0.3 apresentam os diagramas de dispersão da demanda mensal de energia elétrica para as classes residencial, comercial e industrial, respectivamente, e as variáveis exógenas.

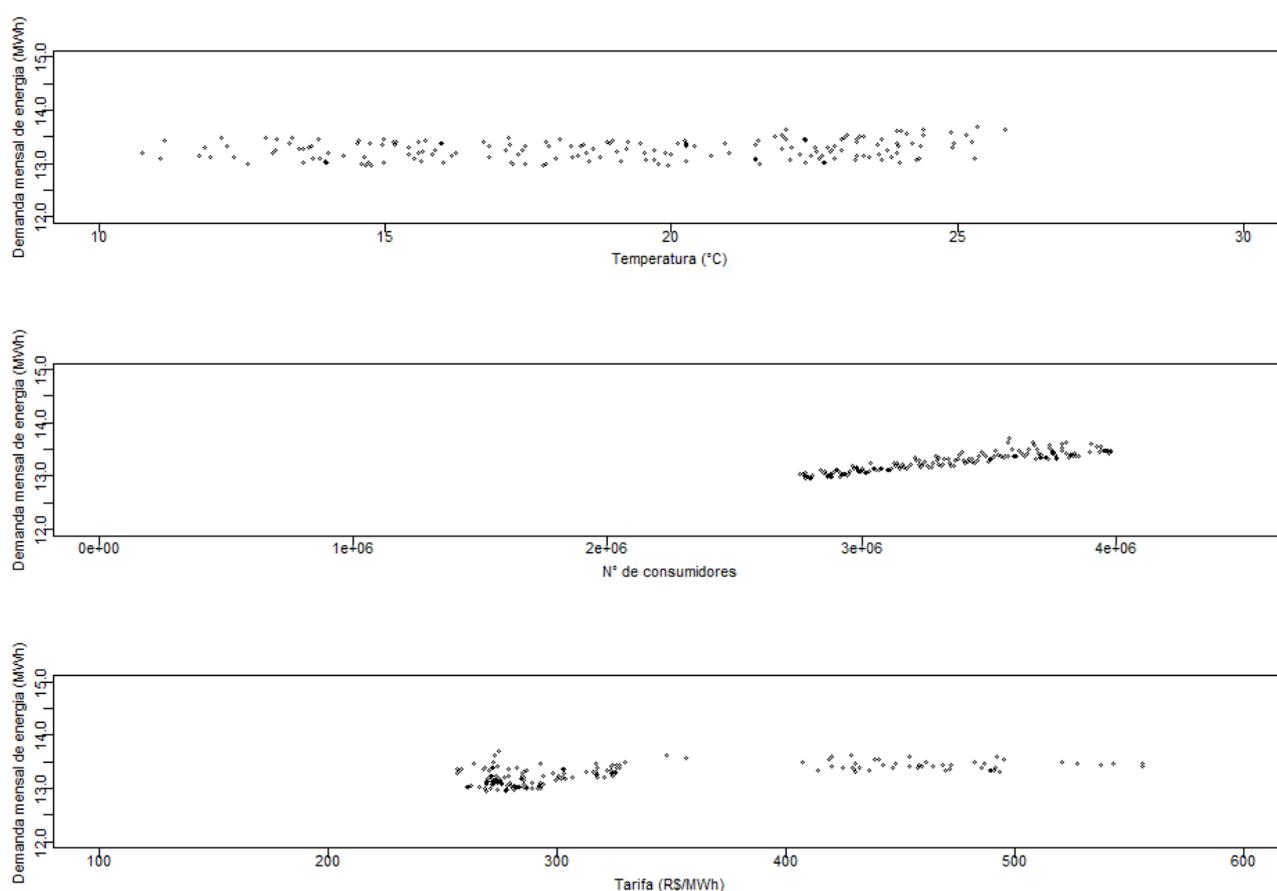


Figura 0.1- Diagrama de dispersão da demanda mensal de energia elétrica da classe residencial com as variáveis exógenas/RS, 2004 - 18

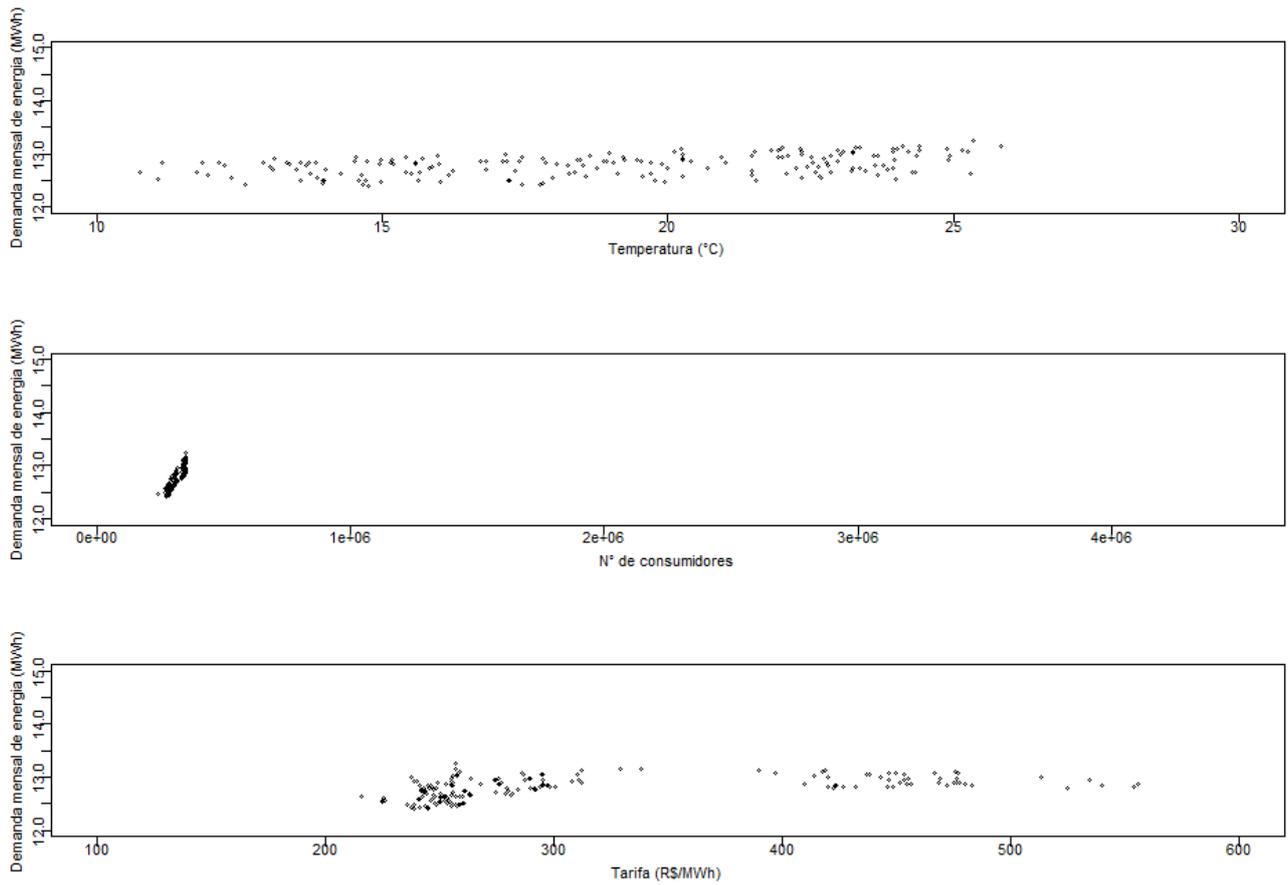


Figura 0.2 - Diagrama de dispersão da demanda mensal de energia elétrica da classe comercial com as variáveis exógenas/R\$, 2004 - 18

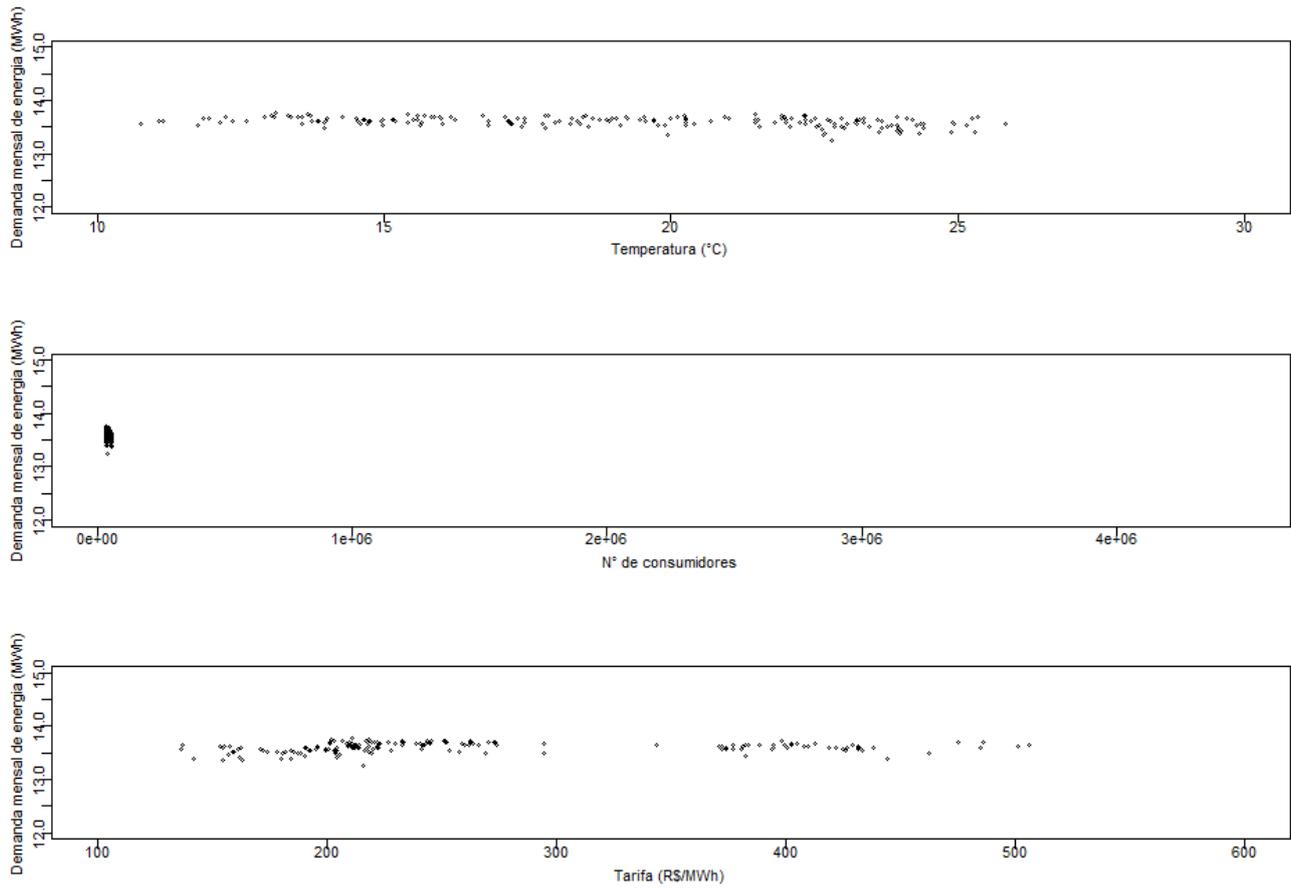


Figura 0.3 - Diagrama de dispersão da demanda mensal de energia elétrica da classe industrial com as variáveis exógenas/RS, 2004 - 18

APÊNDICE 2 – Análise de estacionariedade das variáveis externas

A Tab. 0.1 apresenta o resultado do teste ADF das variáveis externas. Os valores apresentados para avaliação da significância dos coeficientes δ , β_1 e β_2 , de acordo com as características de cada série, indicam a rejeição da hipótese nula para o teste, dando evidências de que a série seja estacionária conforme a ordem de integração apresentada.

Tabela 0.1 - Resultado do teste de estacionariedade para as variáveis exógenas/RS, 2004 - 18

Variável		Parte não sazonal			Parte sazonal			(d, D)
		$\delta (\tau)$	$\beta_1(\varphi_1, \varphi_2)$	$\beta_2(\varphi_3)$	$\delta (\tau)$	$\beta_1(\varphi_1, \varphi_2)$	$\beta_2(\varphi_3)$	
NC _{res}	Crít.	-3,43	4,75	6,49	-3,43	4,75	6,49	(1, 0)
	Calc.	-12,67	53,54	80,31	-4,87	7,92	11,87	
NC _{con}	Crít.	-3,43	4,75	6,49	-3,43	4,75	6,49	(1,1)
	Calc.	-7,99	21,29	31,92	-7,66	-19,56	29,34	
NC _{ind}	Crít.	-3,43	4,75	6,49	-2,88	4,63		(1,0)
	Calc.	-7,02	16,44	24,66	-3,17	5,02		
Tar _{res}	Crít.	-3,43	4,75	6,49	-3,43	4,75	6,49	(1, 0)
	Calc.	-6,77	15,32	22,96	-4,40	6,48	9,71	
Tar _{com}	Crít.	-3,43	4,75	6,49	-3,43	4,75	6,49	(1, 0)
	Calc.	-6,58	14,45	21,66	-3,89	5,05	7,57	
Tar _{ind}	Crít.	-3,43	4,75	6,49	-1,95			(1,0)
	Calc.	-6,86	15,73	23,58	-2,37			
Temp	Crít.	-3,43	4,75	6,49	-3,43	4,75	6,49	(0, 0)
	Calc.	-10,37	35,86	53,77	-6,94	16,10	24,13	

Fonte: Elaborada pela autora